1974

The Explanation Of Human Behaviour In Terms Of Its Rationality

Marilyn Ann Friedman

THE EXPLANATION OF HUMAN BEHAVIOUR IN TERMS OF ITS RATIONALITY

by

Marilyn Ann Friedman

Department of Philosophy

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
February 1974

ABSTRACT

This dissertation grows out of an attempt to reconcile two basic themes with regard to human behaviour: one, the apparent moral relevance which is attributed to human behaviour in everyday life, and two, the equally apparent knowability of human behaviour, in the sense that human actions can often be explained, occasionally predicted, and not infrequently controlled. The former theme suggests the plausibility of supposing that human behaviour is legitimately thought of in terms of such categories as that of free will and responsibility; the latter suggests the plausibility of supposing that human behaviour meets the criteria of knowability (in the sense in which this involves an ability to "give an account of" what is known), in particular, the criterion of lawful relationship, or determination.

Both themes seem equally compelling; yet a tradition of philosophical thought has regarded them as being of at least superficial incompatibility, so that many of those who have endorsed the mutual compatibility of these features have still felt compelled to present explicit arguments to this effect. For reasons which are discussed in chapter four, the usual compatibility arguments of "soft"

determinism are rejected as inadequate, particularly in failing to demonstrate that the full significance of categories of responsibility can be grounded in a metaphysics of causal determinism.

Another form of reconciliation is explored in this work, one which is based on the supposition that the significance of categories of free will and responsibility, while not contra-deterministic, is nevertheless contra-causal. Causation, as a metaphysical relationship of productive activity, is distinguished from determination, as a purely epistemological relationship. In the light of this distinction, it is possible to regard overtly non-causal concepts, such as that of rational self-determination, as comprising an adequate explanatory framework, in terms of logical and epistemological criteria of adequacy, for the explanation and understanding of human behaviour.

The main category of explanatory force, in the non-causal framework for understanding human behaviour outlined in this dissertation, is that of rationality. The attribution of responsibility to an entity for its behaviour is interpreted as presupposing the entity to be a distinct "self" or agent which acts as at least a partial origin of its own behaviour, and which acts intrinsically rationally, i.e. "for reasons". The appropriateness of an action, as expressed by a type of empirical generalization called a "principle of action", is construed as explanatorily

relevant and fundamental in providing an account of why that action actually occurred.

The adequacy of this framework as a reflection of the "true nature" of human behaviour can only be assessed in terms of the long-range success or failure of its application to the study and control of human behaviour. Nevertheless it seems possible to conclude, at this time, that it constitutes a viable and epistemologically acceptable working theoretical framework for the understanding of human behaviour.

ACKNOWLEDGEMENTS

My work on this dissertation has benefited enormously from the helpful encouragement and critical evaluation of Professor Robert W. Binkley, as well as from the help of Professor C. A. Hooker and Professor James J. Leach. Numerous difficulties in the text have been avoided or eliminated as a result of their advice, and any remaining difficulties persist solely as a result of my obstinacy. Work on this dissertation, as a part of my work in the doctoral programme, was generously supported by the Canada Council and by the Department of Philosophy; these sources are gratefully acknowledged. The typing assistance of both Miss Jane O'Brien and Mrs. James Insell has been uniformly excellent. And finally, to Philip, in words that can never be adequate to the task, an expression of my deepest awareness of the unending encouragement, assistance, and trust that have helped to make this project, like our life together, a dream fulfilled.

TABLE OF CONTENTS

covering law model of explanation, my objective shall be to show that the form of understanding and explaining human behaviour which I advocate is as adequate, when judged in terms of the logical and epistemological criteria of this abstract model, as any explanation given in the physical sciences.

III

Ordinary everyday explanations--the explanations of common sense--are usually brief statements the point of which is to provide a piece of information which was previously missing in someone's knowledge of a situation, and without which that previously incomplete account was in some way perplexing or inadequate as an account of the situation. It seems generally accurate, although not very enlightening philosophically, to say that an explanation is regarded as adequate from the point of view of common sense so long as it supplies just enough information to "complete" an account of some explanandum, and thereby render that explanandum "understandable". The factors cited as the "explanation" are frequently brought forward in contexts in which both the person requiring the explanation and the person providing it share a good deal of background knowledge and information. In addition, both parties are frequently aware of this common fund of belief, at least to the extent of taking these shared beliefs for granted, in that they are never

THE STRUCTURE OF RATIONAL EXPLANATION

The scientific study of human behaviour is at least a conceivable endeavour. No ground, logical, conceptual or otherwise, for asserting a priori the impossibility of its success presents itself. This is not yet to prejudge just what the results of such a study will reveal, nor to assert that anything of scientific value is discoverable in human behaviour at all. But so long as human behaviour is counted among observable phenomena, then it must be granted that scientific methods of thought and study can at least be applied to this realm of phenomena, however fruitful or fruitless the results of such application shall prove to be. It may be that the scientific study of human behaviour will ignore or overlook various aspects of human behaviour to which we prescientifically pay regard. But this does not entail that the scientific study, as such, is a totally unacceptable viewpoint from which to study human behaviour, or that it will necessarily result in an account of human behaviour which will be in no way adequate to the subject matter.

Borrowing from Ernest Nagel,[1] I shall characterize a study as scientific if it conforms to the following

characterization: (1) it possesses greater organization and classification than the beliefs which comprise the "Common Sense" of any subject matter; the organization and classification which is important is that which is based upon explanatory principles and aimed at explanations which are systematic and controlled by factual evidence; (2) it aims at removal of the incompleteness of common sense beliefs, particularly the lack of awareness of the limits within which its beliefs are valid and its practices successful; (3) it seeks to eliminate the glaring inconsistencies within sets of common sense beliefs, inconsistencies which are the results of the almost exclusive common sense preoccupation with the immediate consequences and qualities of the observed events that, in a practical sense, human beings happen to value; (4) it seeks greater determinacy and precision of language than common sense, in order to better build the logically integrated systems of explanation which are sought; (5) it neglects the immediate values of things in favour of special sorts of abstract concepts which more readily facilitate systematic and comprehensive explanations and which formulate pervasive structural properties of the subject matter manifested usually only under complex experimental procedures; and (6) contrary to common sense, it subjects its beliefs ". . . as a matter of established principle, to systematic scrutiny in the light of data secured for the sake of determining the accuracy

of those beliefs and the range of their validity".[2]

To sum up, we may thus regard the science of some subject matter as a more organized and systematized set of beliefs on the subject than is held merely commonsensically, one which involves greater precision of language and abstraction of concepts for the sake of systematic explanations of the subject, and which seeks to know the limits of the applicability of common sense beliefs and practices, and to eliminate the inconsistent and unsystematic nature of common sense beliefs, by a procedure of continual testing of its beliefs through refined observational and (where applicable) experimental procedures. As the science of a subject matter transcends the common sense beliefs which pertain to that subject matter, so the scientific study of a subject matter may be regarded as the effort to transcend the corresponding common sense beliefs in the hope of achieving the more complete, consistent, precise, abstractive, tested, and systematized set of beliefs which define the science of that subject.

Granted that the distinction between science and common sense is a difference in degree rather than in kind, we may still recognise that the relatively greater degree of completeness, etc. of science is capable of becoming (relatively) very large indeed, so that the benefits of such study may be potentially very great. What is required of any given subject matter in order that its scientific study

be profitable is only that it not yet be known as fully and adequately as is possible, in accord with the above criteria, on the basis of available common sense beliefs. Of course the results of the application of the scientific method provide the best indication of whether or not such a study is valuable in any given case.

There seems to be no a priori reason to prejudge its profitlessness in the case of human behaviour. Whether or not human behaviour can be exhaustively known by this method, and by this method exclusively is not a question which need concern us. Remember that the point of scientific study has been characterized as the search for only relatively greater degrees of completeness, consistency, precision, abstractiveness, test, and systematization than are available in the current stock of common sense beliefs. We need only regard ourselves as, from the scientific point of view, seeking to improve the "common sense" of human behaviour, and, for this end, to devote ourselves specifically and self-critically to this pursuit for its own sake. There is no reason a priori to think that we are limited to only such "pure" knowledge about human behaviour as common sense derives as a secondary offshoot from the immediate practical concerns of everyday life which are the primary focus of common sense.

Thus I regard the scientific study of human behaviour as a worthwhile endeavour. Additionally, and most

important for the present context, I construe this study as involving the successive epistemic refinement and systematization of the common sense beliefs about human behaviour which preceded it in the evolution of our knowledge on the subject. And in light of the familiar dictum that the social sciences are still in their infancy, it is wise to think that the beliefs which comprise the humble common sense origins of social science are still a very pervasive part of the new "scientific" view of human behaviour, that even within the realm of the scientific study of man we have by and large not yet risen above the common sense ground upon which the science is to be built. There is good reason to regard the common sense view of human behaviour as a still very fruitful source of concepts, hypotheses, and general principles regarding the phenomena of human behaviour which may then be tested and systematized by the methods of science.

In accord with this consideration I shall outline a conceptual framework pertaining to human behaviour which I take to be implicit in the common sense understanding of human behaviour, and which seems to provide a useful and compelling point of departure for the systematizing and refining procedures of scientific study. I characterize this common sense framework for the understanding of human behaviour as "the understanding of human behaviour in terms of its rationality". Human behaviour understood in this

way is, like most phenomena conceptualized from the point of view of common sense, being understood in terms of gross overall macroscopic properties, events, states, dispositions, and sequences--the macroscopic level of categories into which we fit the sequences of human behaviour which comprise everyday life. This level of abstraction shall form the focal point of attention throughout this dissertation. In the remaining portions of the present chapter, I shall first superficially clarify "the understanding of human behaviour in terms of its rationality" and the nature of explanations of behaviour within the context of this conceptual framework, relying upon the intuition of the reader as support for this characterization of the common sense understanding of human behaviour. Subsequently, I shall draw a quasi-formal schema which I think best represents the underlying structure of the various explanations of human behaviour which seem to arise within this conceptual framework.

II

For the purposes of this dissertation, it shall be assumed that accounts which are to constitute explanations must be, at least implicitly, generalizable. Thus, with certain qualifications noted below, the "covering law" model shall form the model of explanation to which I shall endeavour to fit a certain account of human behaviour. My

acceptance of the covering law model has, at once, metaphysical and epistemological dimensions. Metaphysically, I am assuming that something can at most be described, never "explained", if it does not form part of a sequence which exhibits a potentially recurrent pattern elsewhere and elsewhen. Epistemologically, I am assuming that the special intellectual satisfaction derivable from what is called "explanation" is only legitimately realized when that which is to be explained is recognized as being incorporated into potentially recurrent patterns which are apprehended as such.

The implicit generalizability of explanations is frequently called into question in regard to accounts of the explanation of human behaviour, especially in conjunction with a construal of such explanation as being significantly different in character from explanation in physical science.[3] Although it shall presently be argued that sequences of human behaviour, in certain ontological respects, may plausibly and usefully be regarded as significantly different in character from the sequences of phenomena which comprise the subject matter of physical science, it will nevertheless be insisted upon that these differences are to be regarded as ontological, or metaphysical exclusively, and as having no relevance to the logical and epistemological structure of explanations for either sort of subject matter. On all matters of logical and epistemological structure, the

explanation of human behaviour is herein to be regarded as exactly identical to explanation in physical science. And the Hempelian "covering law" model shall be regarded as an idealized abstract representation of this structure in either case.

This assumption requires some defense, especially in light of plausible considerations which have been raised against the "covering law" model.[4] To regard explanations as being implicitly generalizable is, in effect, to regard explanation as requiring subsumption of the explanandum under law. But it is important to recognize that the law-governed structure of explanation need only be <u>implicit</u> in the giving of actual explanations; the factors adduced as explanatorily relevant need not include an explicit statement of the pertinent law, or laws. The point of giving an explanation need not be to <u>show</u> a logical relationship to obtain between a statement of the explanandum and statements of the conditions of the explanans taken as logically antecedent. It is sufficient for the purpose of adequately explaining something that this relationship of explanans to explanandum be implicitly presupposed in the giving of the explanation, which itself need then only <u>explicitly</u> consist of the statement of those conditions which are regarded as explanatorily relevant.

For example, suppose that I explain the blowing of the fuse by citing the fact that you just turned on the

table saw at the same time that the compressor motor of the air conditioner and the compressor motor of the refrigerator were both going on. It is not my intention to show how the blowing of the fuse is a type of event the occurrence of which can be subsumed under some general law. It is not my intention to exhibit a logical relationship between a statement asserting the blowing of a fuse and anything else. My intention is simply to explain why the fuse blew. But the accomplishment of this intention seems implicitly to require that whatever sorts of factors be adduced as explanatorily relevant have a regular relationship of some sort with the type of phenomena being explained. And this regular relationship is implicitly presupposed, and hence implied, by the citing of factors as explanatorily relevant to some explanandum.

The regularity of the relationship between the explanatory factors of the explanans and the explanandum seems important for insuring an adequate explanation on the following grounds. Only our appreciation of the regularity of relationship among types of phenomena could give us any theoretical or conceptual assurance that a perceived sequence of phenomena of instances of those types was a sequence which involved any important unity as a sequence, for example, causal connection. And only if the sequence is thus unified would the occurrence of some components of the sequence have any direct bearing on the occurrence of

other components of the sequence.

This is not to deny that in the fullest sense, each sequence is unique. But if we do not regard connections such as that of causality or determination as numbering among the contents of experience, then our grounds for regarding unique individual cases in terms of such connections seem to have to consist in intellectual readiness to grasp the unique cases in terms of the general patterns which may be abstractly discerned in them. It is not important that we may never again actually re-apply to another sequence a complex pattern discerned in some one particular case that happens to interest us. Explanations of human behaviour frequently involve us in the discerning of a great many explanatory factors which, when finally combined into an explanatory pattern, seem to be, for all practical purposes, non-repeatable as such a combination. What is important if we are to regard such a combination of factors as being conjointly explanatorily relevant in even a single case, is that we are thereby exhibiting our own readiness to <u>infer</u> similarly, should there possibly, however implausibly, recur such a similar concatenation of circumstances at any other time (with due allowance for the intrusion at that time of additional explanatorily relevant factors). Were we not ready to infer in this manner, it would be difficult to determine in just what consisted our acceptance of that concatenation of circumstances as

"explaining" the given explanandum.

"Explanatory relevance" is not being construed in a manner so narrow as to exclude any but causal or mechanistic relationships from satisfying this description. The factor cited as explanatory in a given case may well consist in a reason or justification of the explanandum. The relationship between an explanans of this sort and its explanandum requires philosophical clarification of its own in order that the explanatory relevance of the factor thus related to the explanandum can be understood. An attempt at such clarification is indeed a part of the present dissertation. But the "covering law" model of explanation, in principle, in no way precludes the possibility of such different varieties of explanatory relevance. "Explanatory relevance" is thus, under the present construal, an aspect of explanation which exceeds the logical and epistemological structure of explanation itself and pertains to the way the world is, shall we say, "held together"--the way in which things in the world mutually appertain. The covering law model requires only regularity of explanatorily relevant relationship and does not preclude the possibility that a wide variety of basic types of relationships among phenomena or among other factors will bear the needed regularity.

It bears reiterating, and elaborating upon, that the regular relationship being presupposed in the citing of factors as explanatorily relevant to some given explanandum

need not be given explicit formulation and expression as a universal (or probabilistic) statement in the actual articulation of the explanation. Indeed its explicit formulation may seem trivial and unnecessary once the (logically) antecedent conditions of the explanans have been cited, because the law statement now provides no new information that was not already implied in the very citing of those antecedent conditions. For if the stating of certain conditions as an explanation implicitly presupposes the regular association of those types of factors regarded as explanatorily relevant with the type of explanandum under consideration, then the explicit statement of the law which is presupposed as "covering" the relevant sequence can only trivially and redundantly restate the implication which is obvious in the citing of those explanatory conditions in the first place.

In the context of explaining, say, the blowing of a particular fuse, it will usually be regarded as sufficient if I cite the appliances the use of which caused an excessive surge of electric current to flow through the circuit. I do not need to append to my "explanation" the recitation of some such generality as that whenever the compressor motors of a refrigerator and an air conditioner are simultaneously starting to draw current on a line with a fifteen amp fuse in it, if a table saw is also turned on at that moment then the fuse will blow. Under normal (i.e. non-

philosophical) circumstances, it is likely that such a generality will never as such enter my mind, although anyone who knows about the operation of electrical appliances and about fuses would probably assent to its truth should it ever be stated. More likely, I would probably have conceived the more inclusive generalization that whenever a surge of current through an electrical circuit exceeds the amperage capacity of a fuse on the line, the fuse will "blow". There are a variety of combinations of appliances that can draw such excessive surges of current for fuses of any given amperage capacity and it is unlikely that I shall ever explicitly formulate all the lower-level generalizations to cover each of these combinations. But this fact about the economy of my thoughts on the subject does not entail that there is no generality operative in my explanation of the fuse's blowing on a given occasion.

A different sort of problem for the covering law model arises in the case of explanations which involve appeal to a wide variety of kinds of explanatorily relevant factors. The strongest case for the covering law model arises with explanations in which only one condition or type of condition, or some very small set of different types of conditions, are cited as explanatorily relevant. In such cases, the covering law implicit in the explanation need not be far removed from a statement of the form "If [antecedent conditions], then [explanandum]", and a

fairly tight logical connection binds the whole argument together. Most generalities that we ever have occasion to formulate are economical in form, linking a rather small number of types of antecedent conditions to a type of consequent. Explanations of human behaviour and of social situations, however, seem, in the more interesting and complex cases, to involve us in the adducing of a wide variety of different types of explanatory factors whose mutual reconciliation on a given occasion, in terms of priority of importance and mutual interference, is not a matter for which further generalizations are always available. An ad hoc decision is required to establish the explanatorily adequate combination of conditions for each new case, and no detail is a priori unimportant in this judgment. Clearly there is more structure to the dynamics of explanation in such cases than the simple covering law model reveals.

It seems reasonable to regard the covering law model, based upon the conception of a tight logical relationship between the appropriate covering law and the statements of (logically) antecedent conditions, on the one hand, and the statement of the explanandum, on the other, as an idealized philosophical abstraction. It is a useful and relatively accurate model of the structure of reasoning in the simplest cases of explanation; but it is a model of a structure which must be regarded as having to be compounded

and possibly modified in more complex cases of explanation, in ways which are not themselves determined by the nature and concept of this idealized abstraction. We can regard laws as being of either abbreviated form, expressing simplified abstract relationships among selected variables; or we can regard them as being, even in practice, actually unlimited in the extent of types of variables which they can encompass. Generalities which we commonly have occasion to express are of the abbreviated or economical variety. If we choose to regard the "covering laws" of our model as being, for all practical purposes, limited to this form, then the explanation of a complex event will have to be understood as involving the invocation of a plurality of covering laws for which there is likely to exist no further law which directs the weight and priority to be given to each. This point seems equally true whether the explanandum consists in human behaviour or in physical phenomena. In many explanations of particular complex physical phenomena, a single covering law pertaining to the unique case seems formulable only at the expense of true practical generality--that is, at the expense of economical laws which can really, and not just in principle, be applied to an interestingly sufficient number of additional cases.

For example, if I wish to explain the onset of a particular disease in a particular organism at a particular time, I may have to discuss a variety of factors including

the nature of a virus, the characteristic mechanism by which it is transmitted, the state of various internal mechanisms on the part of the now-diseased organism which ordinarily resist such infection, the environmental conditions which promote the virus' activity, and so on. No known statement of real practical generality encapsulates and mutually relates all these factors to their joint effect upon an organism, although there are likely to be formulated generalizations which relate each factor separately, and thus in abstraction, to the potential of the organism to become diseased. It is a more plausible account of our approach to the practical problem of explanation to think that we keep the simpler more abstractive laws separate in our thinking, for the sake of their wider applicability, and the economy of thinking which they allow. We are thus led to regard the matter of their application, in those cases in which a great many explanatorily relevant factors must be mutually balanced and interrelated, as a process of pure judgment for which no known rules are pre-determined as the correct rules and for which the art of judgment is therefore required.[5]

The covering law model of a simplified explanatory argument does not reveal the structure of this sort of judgment which plays so essential a part in many important explanations. But this is not to deny the philosophical usefulness of the covering law model. Indeed the covering

law model of explanation is a viable account of explanation precisely because of its own economy of thought and wide applicability as an intellectually manageable abstraction. In the more complex explanations in which there is difficulty discerning so simplified a structure of reasoning, there still seems to be intuitive ground for regarding the rationale of the model to be relevant: namely, that generality of some degree is involved in explanation. And this rationale seems to be intuitively reasonable on the grounds that since explanatorily relevant connection (causation, justification, determination) is not to be discovered as numbering among the contents of experience, we seem to have no other experiential feature at which to locate our assurance that various factors are explanatorily relevant to each other beside the expectation that factors of one type are regularly related in some important fashion to factors of another type. Having assented, finally, to some particular explanation of some actual event, through a process of reasoning involving the complex adjustment and reconciliation of many relevant simpler "covering laws", we should be prepared, in all consistency, to explain in similar fashion, further actual events which, in all their uniqueness and particularity, are sufficiently like the event at hand in all the respects which we have judged to be explanatorily relevant this time. This preparedness seems implicit in our very acceptance of a set of factors

as constituting an adequate explanation of a given explanandum. And the philosophical point is in no way affected by the fact that we may never again encounter a sufficiently similar situation, or the fact that we may not expect to encounter ever again a sufficiently similar situation, or the fact that we may never take the trouble to explicitly formulate to ourselves the final complex concatenated generalization expressing the relationship between the mutually adjusted and reconciled explanatory factors and the explanandum. Indeed we should find such a labour trivial and unnecessarily redundant for, as stated earlier, such a generalization is already presupposed in, and hence implied by, our very assertion of the selected factors as explanatorily relevant.

In this account, the covering laws of the covering law model are being construed in a fashion similar to that in which "principles of inference" have been discussed.[6] A full treatment of this view is beyond the scope of this dissertation. Consequently I shall retain the terminology of the original Hempelian model, with the proviso that this terminology be understood in light of the qualifications expressed above. The covering law model shall thus constitute the working epistemological abstraction into which I shall endeavour to fit the account of human behaviour to be outlined in subsequent pages. Thus, without my having to conclusively demonstrate the accuracy or inaccuracy of the

covering law model of explanation, my objective shall be to show that the form of understanding and explaining human behaviour which I advocate is as adequate, when judged in terms of the logical and epistemological criteria of this abstract model, as any explanation given in the physical sciences.

III

Ordinary everyday explanations--the explanations of common sense--are usually brief statements the point of which is to provide a piece of information which was previously missing in someone's knowledge of a situation, and without which that previously incomplete account was in some way perplexing or inadequate as an account of the situation. It seems generally accurate, although not very enlightening philosophically, to say that an explanation is regarded as adequate from the point of view of common sense so long as it supplies just enough information to "complete" an account of some explanandum, and thereby render that explanandum "understandable". The factors cited as the "explanation" are frequently brought forward in contexts in which both the person requiring the explanation and the person providing it share a good deal of background knowledge and information. In addition, both parties are frequently aware of this common fund of belief, at least to the extent of taking these shared beliefs for granted, in that they are never

explicitly cited or invoked in the dialogue exchanged between the two parties.

Given such silence on the parts of persons providing and accepting explanations, what reason is there to suppose that there really may be further elements involved in explanation than those elements expressed by the explicit statements which form the actual explanatory dialogue? Basically we begin by recognizing that many explanatory dialogues are difficult to regard as satisfactory from points of view external to that which the participants share. Statements explicitly offered as explanations may, on occasion, not be recognizable as constituting an adequate explanation of some given explanandum. The account of the explanandum is not yet "complete" from this external point of view, and something important about it may still be puzzling.

For example, suppose that we are puzzled by the fact that at an ordinary room temperature, say 70°F., in a climate like that of London, Ontario, people will generally feel colder in the winter time than they do in the summer. This fact can be explained by citing the much lower relative humidity of the air in the winter than in the summer. Such an explanation, however, will seem quite unsatisfactory as it stands unless it is already known that the rate of evaporation of moisture through human skin is to some extent inversely related to the relative humidity of the atmosphere,

and furthermore that the evaporation of moisture through the skin has the effect of cooling the surface of the body. Once supplemented by these general statements, the explanation given above seems quite adequate. If such information is already a part of one's standing set of background beliefs, then the initial statement citing the relative difference in humidity between summer and winter should have been sufficient in itself to have satisfied any request for explanation of the phenomenon. In that case the background beliefs would have provided a context of explanation enabling a proper and ready appreciation of the explanatory significance of the single factor initially cited.

We can recognize that even in cases where we ourselves rely upon background belief to provide a context for the ready appreciation of abbreviated explanations that we offer, the "full" or "complete" statement of the explanation would require explicit expression of these presuppositions. On many occasions we just do not bother to recite such information as part of our explanations because we can afford to take for granted a sharing of those same background beliefs on the part of our audience. Only when the audience continues to be puzzled about some explanandum despite the statement of some such abbreviated explanation must we fall back upon a little self-conscious reflection in order to discover what else must be added to the original explanatory statement in order to allay the perplexity.

Ordinarily we can, if we are careful and patient, discern the presuppositions of our thinking and give them an explicit formulation if required.

In outlining the structure of common sense reasoning about human behaviour, I shall attempt to elicit as fully as possible the different sorts of beliefs which tend to be presupposed, and hence not stated explicitly, in the giving of ordinary common sense explanations of human behaviour. This effort to achieve as complete an account as possible of the explanation of human behaviour in terms of its rationality should occasion no philosophical perplexity. What is noteworthy about it are the different degrees of generality of the background beliefs revealed by self-conscious reflection. Of particular interest and importance is the background belief which is not of a direct "factual" nature at all, but is more on the level of a "framework principle", definitive of the entire conceptual framework of human behaviour that I shall be exploring. Essentially this is the presupposition that human beings act rationally. This presupposition will be explicated and defended in subsequent chapters.

The criterion for assessing the "completeness" of an explanation, that is, its completeness when account is taken indiscriminately of what is explicitly stated on a given occasion and what is silently presupposed, can only consist in some degree of logical adequacy. The explanans

must, to some minimal degree, show its deductively entailment of the explanandum. On this ground, it does not matter what has been explicitly stated and what has been silently presupposed, and among the latter it does not matter whether the presupposition is of the nature of factual belief or "framework principle". However there is reason for maintaining a working distinction between framework principles and other presupposed beliefs which enter the completed explanations of human behaviour. By depicting the form of explanation of human behaviour which I am defending in a manner which more naturally exhibits the different levels of our thinking in the actual giving of such ordinary common sense explanations, I am more likely to elicit intuitive agreement from the reader as to the accuracy of this account as a structural account of such common sense explanations. Nevertheless logical adequacy still remains the criterion by which the completeness of such explanations, as arguments, is finally to be judged.

IV

My objective is to depict a certain form of understanding of human behaviour which I believe to characterize the common sense view of human behaviour. When fully sketched out, this form of understanding is as adequate, in terms of the logical and epistemological criteria of the covering law model, as any understanding which is achieved

in the physical sciences. However, as I have pointed out, the covering law model, in regard to any domain of knowledge or belief, is most clearly applicable to relatively simple cases of explanatory argument, explanations in which only one or a very few explanatory factors are invoked. Consequently my efforts will be made easier at the outset if I concentrate on relatively simple cases of the explanation of human behaviour in terms of its rationality. The intuitive plausibility of this form of understanding of human behaviour will also be more apparent to the reader in regard to such homely, uncomplicated cases. I shall therefore take the liberty of restricting attention to such cases for the present.

Suppose that explanation is requested for the fact that Smith is taking two aspirin. Suppose that, in response, he says that he has a headache. Such a response would normally be regarded as quite adequate. Nothing puzzling remains in the facts of the case as stated, and under normal circumstances there is probably no additional factor in the situation that need be explicitly cited in order that Smith's action be "explained" to the satisfaction of whoever has requested the explanation. Were we to begin to schematize the "explanation" which we have so far elicited, it would look something like this:

Smith has a headache.
∴ Smith takes aspirin.

In most ordinary cases in which the taking of aspirin is explained, the explanans above would be regarded as sufficient as it stands. But clearly the example has yet to meet the requirements of logical adequacy.

Our ordinary appreciation of such an explanans as adequate occurs within the context of certain common and pertinent background presuppositions. Among such relevant background beliefs, might be the view that for most people, the ingesting of aspirin can eliminate many sorts of mild to moderate bodily pains, especially headaches. Consequently, it can be assumed that when one has a headache, then an appropriate thing to do is to take aspirin. The reliance upon such an assumption in a given case may require further background presuppositions. For instance, it is not "appropriate" to take aspirin for a headache if one is affected by aspirin in such a way that a reaction will occur which is more painful than the headache which the aspirin eliminates.

With due regard for such complicating circumstances, there are still two senses in which it may be correct to say that it is appropriate to take aspirin for a headache. Firstly, one may disregard all additional factors and express an abstracted relationship between aspirin and headaches which thus obtains <u>ceteris paribus</u>, or with all other factors being momentarily regarded as irrelevant to this relationship. This abstract generality is doubtless

one which most of us accept in our daily lives, and which we tend to regard as applicable to headache-situations unless positive interfering factors in particular situations prove its inapplicability. Secondly, we may encompass all additional relevant variables in a given headache-situation, for instance in Smith's case, and on the basis of this express the final considered judgment that (in this particular case) the taking of aspirin would be an appropriate way (perhaps the appropriate way) to cope with a headache. This judgment would rest upon a recognition that all relevant variables would possess values in the given case which do not in fact interfere with the effect of aspirin upon headaches as this relationship is conceived in abstraction from these variables.

The latter judgment would have embodied a process of explanatory dynamics that we have found to be totally ignored by the covering law model. For, in effect, a plurality of simpler laws, each having great practical generality, has been found relevant to a given case, and the final considered judgment expressed above represents the conclusion based upon their mutual reconciliation. This final judgment, although in principle generalizable (once the relevant conditions are explicitly included in its statement), is not a "law" in any practical sense. It would not possess the economy of thought or the real practical generality to operate as a standing background belief which

could efficiently serve in the understanding and explanation of new headache-situations involving new and different combinations of values of relevant variables. Only the former abstract generality--that it is appropriate to take aspirin for a headache, ceteris paribus--can conveniently operate in this capacity. If the covering law model of explanation is really to point to a covering "law" and not a covering tautology, then it would have to be in this latter sense that the notion of the appropriateness of aspirin for a headache is invoked by our acceptance of Smith's response as an adequate explanation for his taking of two aspirin.

Let us therefore expand the original representation of the explanation in question to include a statement of this presupposed background belief:

Smith has a headache.
When one has a headache, an appropriate thing to do is to take aspirin.
∴ Smith takes aspirin.

There is still more that must be added to complete the explanation, not only to satisfy the condition of logical adequacy, but also to provide an explanatory argument from which can be extracted a more generalized explanation schema which covers all the various sorts of (simplified) behavioural explanations that we ordinarily deal with. Not all explanations of human behaviour center around the explicit statement of a bodily condition of the agent, or even, more generally, around the explicit statement of a

circumstance confronting the agent, as "the explanation" of the agent's behaviour.

A variety of factors may be explicitly appealed to in order to explain human behaviour, and against an appropriate set of background beliefs, all such appeals may satisfy the listener and may even constitute adequate explanations on logical grounds--that is, as taken in conjunction with a complete statement of the relevant background beliefs. Though the explicit explanatory statements can thus be of a variety of different sorts of factors, careful consideration and analysis suggests that the entire explanatory accounts may well be identical in structure in all such cases. The "structure" which is of interest here is no longer the simple matter of logical entailment of the requirement that a generalization be somehow involved in the explanans. These aspects of the idealized covering law model of explanation have already been adopted as the working model for an adequate account of the explanation of human behaviour in terms of its rationality. The additional structural matters that are of special interest in the present context pertain to the classes of entities, both cited and presupposed, and the relationships presumed to obtain between these, which together are commonsensically regarded as explanatory of human behaviour. Bearing in mind the previous (unfinished) example of an explanation of behaviour, let us consider another somewhat different sort

of case.

Suppose that Jones arranges to attend a technical high school. If asked why she has made this arrangement, suppose that she replies (truly) that she wants to study drafting. In order to understand the full relevance of this reply, we shall need to know that drafting courses are not generally taught at the ordinary public high schools which are oriented toward preparing students for college. Only a technical high school will provide Jones with the opportunity for training in her chosen career. The explanation, including these pertinent background assumptions, has roughly the following form thus far:

Jones wants to study drafting.
If one wants to study drafting, an appropriate thing to do is to attend a technical high school.

∴ Jones attends a technical high school.

A more abstract schematization of the important structural aspects of such an explanation is the following:

Jones wants E.
If one wants E, an appropriate thing to do is A.

∴ Jones does A.

The types of factors in the situation which are conjointly explanatorily relevant to Jones' action are the want of the agent and the appropriateness conferred on the action to be explained by the existing circumstances, given that want.

Superficially this example may appear to be distinct in form from the previous example. A more abstract schematization of the previous example would be as follows:

Smith is in circumstance C (has a headache).
When one is in circumstance C, an appropriate thing to do is A (take aspirin).
∴ Smith does A.

The important surface difference between the two cases is that in the Jones example, the explicit explanatory factor is the want of the agent, whereas in the Smith example, it is a circumstance which, shall we say, "confronts" the agent. But closer survey of the examples reveals that the same total variety of types of factors actually underlies our appreciation of, and acceptance of, either explanation as an explanation. In Jones' case, the action which is appropriate to achieving Jones' objective is made appropriate by circumstances which confront Jones, for example, the fact that drafting courses are not taught at ordinary high schools. In the absence of such facts as this, it could not in general be said that an appropriate thing to do in order to study drafting is to attend a technical high school. Only the full combination of wanted ends plus circumstances which restrict the means by which such ends may be attained can lead to any conclusions about what are appropriate things to do.

In Smith's case, the same observation holds. The action which is made appropriate by the circumstance

confronting Smith is made appropriate, and not just a possible alternative action by the (presumed) want on Smith's part to be rid of pain. Were we, for some reason, not able to presume this very pervasive human want to characterize Smith, then we could (as yet) have no ground for regarding the general appropriateness of taking aspirin for a headache as having any explanatory relevance to Smith's particular case. Again only the full combination of end plus circumstances relevant to the achieving of this end can lead to any conclusions about what are appropriate things to do in situations characterized by those circumstances.

Thus, the varied enthymematic explanations ordinarily given of human actions seem, in at least a great many cases, to point beyond themselves to a fuller explanatory account which may, so far, be schematized thus:

X wants E.
X is in circumstances C.
If an agent wants E, then if that agent is in circumstances C, an appropriate thing for the agent to do is A.
∴ X does A.

A good deal of explication must be provided for the generalization involved in this form of explanation, namely, "If an agent wants E, then if an agent is in circumstances C, an appropriate thing to do is A". This discussion will be postponed to a subsequent chapter.

At present the objective is to show in schematic outline how the requirements of logical adequacy can be met by common sense explanations of human behaviour, as exemplified above. The important logical gap in the above schematized account which remains to be crossed is the gap between statements of the form "An appropriate thing to do is A" and statements of the form "A is done". There are (at least) two ways in which provision can be made in our explanatory model for the crossing of this gap. The one which we choose to include in the idealized abstract model of the explanation of human behaviour should be the one which we feel, on the basis of self-conscious reflection, to most adequately represent the essential features, explicit and presupposed, which are involved in the everyday attempts to understand why various human actions have occurred.

One way in which the above model of explanation can be modified so as to achieve logical adequacy results in the following schema:

X wants E.
X is in circumstance C.
If an agent wants E, then if that agent is in circumstance C, that agent does A.
∴ X does A.

For my purposes, this schema cannot be said to be incorrect as a representation of the overall structure of the explanation of human behaviour in terms of its rationality. But if we adjust the model in this fashion, or in some

fashion which is comparable in so far as the intermediary reference is dropped to what is the appropriate thing to be done under the given antecedent conditions (end plus circumstances), then the model of explanation no longer reveals whether it is through an intermediate appreciation of the appropriateness of an action under given conditions, or whether through consideration of some other factors, that we finally come to regard the performance of action of the type A as, in general, actually occurring upon the obtainment of the given conditions. It is, however, a distinctive feature of our ordinary understanding of human behaviour, as I interpret it, that the occurrence of an action is regarded as "explained" by the citing of such antecedent conditions of the situation of its occurrence as, together with relevant background information which makes up the context of explanation, show the action to have been appropriate, reasonable, proper, justified, right, suitable, satisfactory--in short, rational, in the situation in which it occurred.

This appreciation of rationality is an essential step in the understanding of those events in the sequences of human behaviour which are distinctively human, and distinctively a matter of behaviour in the fullest sense. Behaviour in the fullest sense, or "action", may be contrasted with physiological or biological or mechanical regularities which can also be discerned in the sequences

of events pertaining to human organisms, but for the explanation of which no appreciation of the appropriateness of the explanandum is typically involved. Again, a good deal of explication is required to defend my views in this regard, and this discussion too will be postponed for the present. The point to be stressed at this time is that in order not to lose any reference, in the final schematization of the common sense form of explaining human behaviour, to the appreciation of the appropriateness of action, regarded as essential to the ordinary understanding of human behaviour, it is necessary to schematize this form of explanation as follows:

X wants E.
X is in circumstance C.
If an agent wants E, then if that agent is in circumstance C, an appropriate thing for the agent to do is A.
People do what is appropriate for them to do.
∴ X does A.

This new generalization which has thus been introduced into the explanation schema is, I shall be arguing, on a different level from that of the generalizations pertaining to the appropriateness of various types of action. This is the standing presupposition which operates as the "framework principle" to which I alluded earlier. In effect, it defines the very structure to which explanations of human behaviour must conform, namely a structure which involves the appreciation of the "appropriateness" of human

action as an integral part of the process of explaining the occurrence of that action.

Since the "appropriateness" of action consists in its efficacy, in given circumstances, in achieving wanted ends, the standing presupposition that people act "appropriately" is a presupposition to the effect that people act so as to achieve the things that they want, i.e., that people act rationally. Such a presupposition is virtually tautologous. Certainly it is never open to direct empirical test in any immediate way. It may, however, be regarded as being tested indirectly through the overall long-range use of the very system it defines for understanding human behaviour.

As a more adequate reflection of these two facts about this generalization, namely, that it operates as a standing framework principle which controls the remaining structural aspects of the explanation of human behaviour, and that its basis of acceptance is other than direct test or any relatively limited set of observations, it seems worthwhile to divide the above schematic representation in the following manner:

People do what is appropriate for them to do (i.e. people act rationally).	X wants E. X is in circumstance E. If an agent wants E, then if an agent is in circumstance C, an appropriate thing for that agent to do is A.

∴ X does A.

A fuller explication and defense of the presupposed "framework principle" will be provided in subsequent chapters. At present I shall merely indicate the sort of intuitive plausibility which it seems to have as an expression of the most general presupposition lying behind our ordinary explanations of human behaviour. Consider the earlier example in which Smith's taking of aspirin was explained by citing Smith's headache. Together with the reasonable presumption that Smith would want to be rid of bodily pain and the presumption about the efficacy of aspirin in accomplishing such wants, we seem to have adequately explained Smith's action. Better yet, have we not frequently encountered actual human actions which were clearly explainable on just the sorts of grounds provided in this case, namely, some wanted end and the circumstances which make a certain action appropriate to achieving it? In most cases of this sort, there seems to be nothing more that needs to be said in order that an "explanation" be taken as accomplished. One important objection which should be noted even at this early stage of analysis, is that we must also presuppose that the agent whose behaviour is being explained is <u>aware</u> of the explanatorily relevant circumstances as cited, and is <u>aware</u> of how they enable the fulfillment of what he wants. This legitimate consideration shall eventually be provided for in the full explication and defense of the above model of explanation which is to follow.

such a claim, for all the progress of the universe through time of which we can each be aware is encompassed by one non-reversible "journey" through a sequence of states no two of which are entirely identical with each other. Were the universe to be cyclical and each of us to fully and completely "live again", then by definition each would again only be aware, the next time round, as of a single non-reversible "journey". So barring such empty possibilities, we must suppose that in the course of the "unfolding" of the universe, no complete state ever occurs more than once.[1]

In that case, the universe as a whole throughout time would be most accurately described by means of an appropriately ordered list of descriptions each exhaustive of some temporally minimum state of the universe. But in accord with such a descriptive list, perhaps infinitely long, each state of the universe would, for all we know, be determined not only by its temporal predecessor, but also its temporal successor, or, for that matter, by any and every other state on the list. For it is not only the immediately preceding state but all the members of the entire sequence as a whole that determines the position of any particular state. Such determination amounts to nothing more than determination, or ordering, of the position of each state in the (one and only) time-series. Such a sense of determination would not entail that there are ongoing natural laws which relate the <u>character</u>, in all or any of

believe, however, that these criticisms can be handled by minor revisions of the model, or by qualifications upon the way in which it is to be understood (to be provided in chapter seven), all of which stops far short of seriously undermining it as an adequate explanatory model. It is fairly certain that "common sense" contains some implicit mechanism for the successful understanding of human behaviour; our frequently correct everyday expectations with regard to the activities of persons whom we know or of public and well-known personalities do not, in most cases, rest upon the application of any refined, scientific procedures. The major difficulty here is to find the correct explanatory model to represent this, at least moderately, adequate form of understanding of human behaviour which clearly is present in our shared fund of everyday beliefs about the world. I hope to demonstrate that the model schematized above is the correct one, and thus constitutes the primary point of departure for refined, systematic scientific exploration of the behaviour of human beings.

FOOTNOTES

[1] The Structure of Science (London, 1961), ch. 1.

[2] Ibid., p. 12.

[3] Cf. William Dray, Laws and Explanation in History (Oxford, 1957), esp chs. 1, 2, 3, 5. Also Richard Taylor, Action and Purpose (Prentice-Hall, 1966), pp. 218-220.

[4] Cf. Dray, chs. 2, 3. The basic points treated in the subsequent discussion are drawn from this work, in which they are raised as criticisms of the covering law model.

[5] Cf. Michael Scriven, "The Key Property of Physical Laws--Inaccuracy", in Herbert Feigl and Grover Maxwell, eds., Current Issues in The Philosophy of Science (New York, 1961), pp. 91-104; and Michael Scriven, "Truisms as the Grounds for Historical Explanations", in Patrick Gardiner, ed., Theories of History (Glencoe, 1959), pp. 443-475.

[6] Cf. Dray, op. cit., pp. 40-44. Gilbert Ryle, "'If', 'So', and 'Because'", in Max Black, ed., Philosophical Analysis (Ithaca, 1950), pp. 323-340.

II

DETERMINISM

In this dissertation, as outlined in the preceding chapter, it shall be argued that much of human behaviour can be understood in terms of lawful patterns which it exemplifies, and which are prima facie rather distinctive. The patterns in question are those which may be said to reveal the "rationality" of behaviour. Essentially that behaviour shall be construed as rational which is performed as a means to achieving wanted ends. Thus, it is to be maintained that human behaviour, at least to the extent to which it exhibits such patterns of rationality, is deterministic. By claiming that human behaviour which is rational is also deterministic, I am claiming that such behaviour is predictable, in principle, and therefore a proper object of scientific investigation and of knowledge in general. However, I hope also to eventually make compelling the view that human behaviour which is rational as well as deterministic may also quite plausibly be regarded as being "freely willed" behaviour.

Usually, when the existence of free will is asserted to be compatible with the truth of the philosophical thesis of determinism, a variety of "soft determinism" is intended. According to such a view, free will is

regarded as obtaining in the realm of human behaviour just in so far as the temporally antecedent conditions which determine behaviour include among their number conditions of certain sorts, for example, beliefs, desires, volitions. And quite frequently such conditions are thereby conceived as causing, in all respects, the behaviour which is thus "freely willed", in the sense of constituting a sufficient set of what I shall call "temporally antecedent causally efficacious and causally necessitating conditions".

My own thesis is not of this sort. Indeed if I had only to advance yet another variety of soft determinism, I should have difficulty in finding anything new to say. By claiming that a human action manifests "freedom of the will", I wish to say rather that, at the least, there is an element of "spontaneity", "originality", "creativity", "novelty", or "initiative" which is to be regarded as being involved in some way in the lawful sequence which comprises the freely willed action and its determining conditions. What I am rejecting is the view that determined behaviour which is also freely willed is also caused in all respects by its determining conditions. I wish, thus, to draw a distinction between the thesis of determinism and a thesis which I shall call causal determinism, and which consists of the deterministic hypothesis formulated in terms of determining conditions construed as sets of individually necessary and jointly sufficient "temporally antecedent

causally efficacious and causally necessitating conditions". It is this latter thesis the truth of which I regard as being incompatible with the existence of free will, and, consequently, it is the latter thesis which I reject as an acceptable construal of the patterns of rationality exemplified by human behaviour.

Few people really deny entirely the existence of patterns in human behaviour and the significant amount of predictability of which human behaviour is possible, even in ordinary everyday contexts which conspicuously lack anything approaching a systematic scientific account of the matter. I am in need of a new winter coat to better withstand the long Canadian winter; I arrange to go to a store and buy one. It is possible that my need might have been met in other ways: I might have stolen or sewn a coat, or continued using my old coat. But stealing involves the risk of being caught, which raises the strong possibility of incurring penalties that are distasteful even to contemplate. And to steal from someone is to do that person a material injury which, so far as I know, the shopkeeper does not merit; so that even failing to be caught, I would have harmed a person who, in respect of his relationship with me, was entirely innocent. Sewing a winter coat requires time and patience, not all of which I possess in any great amount, and what I have can more skillfully and successfully be applied to other ends. As for my old winter coat--well,

so much for this example. The point is that people commonly act as I have done; it is quite a sane and practical way to handle the situation which I now face. Anyone else who acts similarly in a similar situation would be equally considered to have handled a minor everyday problem in a way that is similarly so sane and practical that it is hardly worthy of note. The following sequence is quite understandable: needs new coat + rejects stealing + rejects sewing ⟶ buys coat. We can, if we bother, predict, with accuracy, that people whom we know are going to buy new winter coats. That something in the way of determination obtains in the realm of human behaviour is thus intimated quite strongly by such ordinary everyday accounts. The problem which is at issue in regard to the free will question is not that of whether human behaviour exhibits any regular sequences which can be expressed in terms of abstract generalizations of some kind and can serve as the basis for predicting future instances of human behaviour. What is at issue, as I shall attempt to show in this chapter, is the metaphysical interpretation to be put upon the distinctive sorts of deterministic sequences which human behaviour is generally acknowledged, in one way or another, to exhibit.

II.

The thesis of determinism is an account of the possibility of our attaining certain important epistemological goals which we frequently set for ourselves with regard to the world of our experience: prediction, explanation, the success of the scientific enterprise and the achievement of knowledge in general. Indeed, the thesis itself seems to be an essentially epistemological thesis, in that it most importantly concerns the structure which our thinking must achieve in order that, by so thinking, any of the above goals may be attained. As an essentially epistemological thesis, determinism appears to involve a minimum of ontological commitment. That is, a minimal amount in regard to the structure requisite for being the "object" of knowledge, is postulated by this thesis as obtaining in the world of our experience. In particular, if the account of determinism, which I shall delineate, is correct, then it is the case that determinism does not involve or entail the view that a set of determining conditions amounts as well to a set of (individually necessary and jointly sufficient) temporally antecedent causally efficacious and causally necessitating conditions. That is, determinism does not involve or entail causal determinism. The latter view constitutes a metaphysical interpretation of the relation of determination. It is to this view that the doctrine of free will, in terms of spontaneity,

constitutes a mutually incompatible metaphysical alternative. The conclusion of this chapter will be that, even assuming some form of determinism to be a necessary condition of epistemological achievement, since causal determinism in particular is not a presupposition for the possibility of such achievement, i.e., its truth is not entailed by the fact of such achievement, therefore the doctrine of free will, expressed in terms of spontaneity, is not a priori inconsistent with the possibility of epistemological achievement with regard to (freely willed) behaviour. And if the doctrine of free will, as a metaphysical interpretation of the relation of determination (in the realm of human behaviour), is in fact consistent with the thesis of determinism, then it is a doctrine the truth of which is quite compatible with the successful prediction, explanation, scientific study, and knowledge, in general, of human behaviour.

Suppose that we vaguely and intuitively regard the world as forming a coherent orderly whole. We might attempt to express this view more precisely by saying that each state of the universe as a whole, that is, in all its basic constituents (things, or the characters, positions, or relationr of things), in any minimal time period, is exhaustively determined by the state of the universe as a whole, of the immediately preceding minimal time period. As a matter of fact, there can be no verifiable content to

such a claim, for all the progress of the universe through time of which we can each be aware is encompassed by one non-reversible "journey" through a sequence of states no two of which are entirely identical with each other. Were the universe to be cyclical and each of us to fully and completely "live again", then by definition each would again only be aware, the next time round, as of a single non-reversible "journey". So barring such empty possibilities, we must suppose that in the course of the "unfolding" of the universe, no complete state ever occurs more than once.[1]

In that case, the universe as a whole throughout time would be most accurately described by means of an appropriately ordered list of descriptions each exhaustive of some temporally minimum state of the universe. But in accord with such a descriptive list, perhaps infinitely long, each state of the universe would, for all we know, be determined not only by its temporal predecessor, but also its temporal successor, or, for that matter, by any and every other state on the list. For it is not only the immediately preceding state but all the members of the entire sequence as a whole that determines the position of any particular state. Such determination amounts to nothing more than determination, or ordering, of the position of each state in the (one and only) time-series. Such a sense of determination would not entail that there are ongoing natural laws which relate the <u>character</u>, in all or any of

its respects, of each complete state of the universe to the character of its most immediately preceding or succeeding state.

It is not much help to invoke the existence of some law or function by means of which any state of the universe can be calculated, and perhaps predicted, given the values for some arbitrary initial state.[2] This formulation of the thesis of determinism has barely more empirical significance than the previous formulation. For we are told that mathematicians can construct a formula, though perhaps of a high degree of complexity, which is capable of representing each state of the material universe as a function of time.[3] Again this thesis of determinism would not entail that there are ongoing natural laws which relate the character of any state of the universe to any other. Furthermore, as G. J. Warnock has put it, "A world whose behaviour could only be rightly described in statements of law too complicated for us to formulate or comprehend would be, for our purposes, every bit as intractable as a world whose behaviour was merely random and chaotic—it would in fact be indistinguishable from such a world".[4]

Warnock's comment brings us to what may be characterized as the whole point of the thesis of determinism: namely, that of formulating the conditions under which knowledge of the world is possible. Of course such a thesis is useless to us if it does not formulate the

conditions under which knowledge is possible for human beings; it does not so much matter what Laplace's "Superior Intelligence" is capable of achieving. The possibilities of human knowledge of the world are limited by certain rather obvious circumstances. Direct experience of the entire world is severely restricted for each human being. Thus, if any sort of determination is to make itself manifest to the human mind, it must be of a sort which will be recognized as such within the context of these limitations. In addition, it seems that the only sort of determination which can be recognizable as such within the limitations of human experience is one in which segments of experience--things, qualities, relations, events, states, processes--are grasped as comprising necessarily unified sequences, sequences which are necessarily unified at least in virtue of exhibiting patterns through space and/or time which are potentially recurrent at future times and which may reasonably be grasped as such. For if extrapolation of the postulate of determinism to the whole world is to be empirically legitimate, then the regularity which we shall want to postulate of the whole world must encompass regularities of which we can become aware within that experiential base. That is, we must be able to become aware of necessary, potentially recurrent, unified sequence patterns within, and among segments of, our experience. Finally, it seems to be a guiding presupposition and

rationale behind any formulation of a deterministic thesis, that what is being formulated are the conditions of something which human beings at least occasionally do successfully, namely, acquire knowledge of the "whole world". For it is generally going to be something which human beings at least occasionally do that will be regarded as a reliable indication as to what ought to be done to acquire such knowledge. The existence of at least some human knowledge is thus presupposed as fact.[5]

Several dimensions of limitation upon the nature of theories of deterministic form arise out of the limits upon human experience noted above. Firstly, if such theories are to be supported by some degree of practical confirmation, or at least, by the apprehension that they have some measure of "fit" and consistency with experienced aspects of the world, then such confirmation and fit must be able to "show itself" within highly limited stretches of experience even with regard to theories the generality of which involves generous extrapolation beyond those stretches. The sequence patterns by which such confirmation is achieved must be patterns of relatively short temporal duration and of relatively limited "cross-sectional", or spatial, extent. It must be possible for the sequence to be apprehended as a sequence and not just apprehended sequentially. This point can be put by saying that it is required that laws which are to be of any use to human beings in the knowledge of the world pertain to

limited systems of variables regarded as the relevant variables, all excluded variables regarded as being irrelevant to the patterns which obtain among those selected variables. Only within the context of such highly limited systems would the potential recurrence of patterns become a likely possibility and the actual recurrence of patterns be recognizable as such to human beings.[6] Thus, rather than considering the determination of whole states of the universe by whole states of the universe, we must consider determination of whole states of systems, by whole states of systems.

Secondly, there must be available a method for correlating the precise mathematical values derived from application of a law or function to a given description of a state of a system with the (to some degree) imprecise measured values of the "determined" state of the actual system. Some degree of tolerable margin of difference between law and experience must be acknowledged to be consistent with the possibility of human knowledge based upon that law, if not as a "necessary" fact about human knowledge, then as a contingent fact of which we surely have the greatest practical evidence.

Thus the limitations upon human experience impose limitations upon that which can constitute an adequate theory of experience which is to be deterministic in form. Out of these considerations, a philosophical theory of

determinism of the following sort begins to emerge. As a first approximation to the thesis, we might posit that any "item" in experience can be calculated, in principle, on the basis of some given set of initial conditions (exhaustive description of relevant variables of the state of the selected closed system) together with the law or set of laws descriptive of all relevant changes among those conditions, and calculated with some specifically tolerable margin of possible deviation from measured value. It is common to find the thesis expressed in terms of predictability rather than in terms of calculability,[7] but this manner of formulation seems to involve certain problems. Most generally, there seems to be no basis for assuming, even if an "item" in experience cannot be known except by being grasped as part of some potentially recurrent unified sequence pattern, that, in addition, every such knowable item could in principle have been grasped in this way prior to the time of its occurrence.[8] Of course prediction <u>in fact</u> is one of the most compelling empirical <u>signs</u> of determination to us, and a strong sign furthermore that we have in our intellectual grasp something in the <u>vicinity</u> of whatever laws most adequately express the deterministic relationships of the system at hand. Successful prediction, on at least some occasions, may even be said to constitute one necessary condition for legitimately being able to speak of some set of laws as constituting an adequate

(deterministic) theory with regard to the changes occurring among some specified system of variables. However, if the predictability of each and every "item" in experience were in principle impossible, this would not entail that the thesis of determinism was false though it might entail that experience could not confirm for us the unlimited applicability of the thesis to the world of our experience.

Is relative and limited "confirmation" of this thesis available to us? In one sense, yes; in another sense, no. Experience confirms the thesis of determinism to the extent that genuine knowledge of the world is gained as a result of grasping elements of experience as forming potentially recurrent unified sequence patterns. But it is not clear that there is an independent way of knowing that it is genuine knowledge and not mistaken or unjustified belief which has been gained instead. For the thesis of determinism amounts almost to a definition of knowledge in terms of just this sort of grasp, and to that extent it is not really an empirically verifiable thesis about the conditions of knowledge at all.

But the thesis also involves a further claim about the knowability of all elements of experience. It is thus a thesis not only about how things (which are knowable) are known, but also about the (in principle) knowability of all things (by those methods by which alone things are genuinely known). Limited confirmation of this thesis may seem to

consist in the success, within science, at calculating, by the use of postulated laws, and within presumably tolerable margins of divergence from measured values, the future values of (relevant) state variables of certain systems on the basis of given values of those variables.[13] It seems that the thesis is definitive of an entire conceptual approach to the attempt to gain knowledge of the world, and as such, it is not merely another scientific theory subject to confirmation by the usual methods of scientific experimentation. If anything, the thesis expresses a rationalization of those scientific methods which seemed to "confirm" the thesis, by defining what appears to be a fundamental conceptual orientation of the whole scientific approach. It would seem that the most that can be hoped for with regard to this philosophical thesis is a "fit" or consistency of that approach to those knowledge-seeking and knowledge-applying situations which are successful, in the sense of enhancing survival and enabling the achievement of other such pragmatic goals. Its own "confirmation" in any more practical or active sense could only be inherently circular for it is a theory, at least partly, about what in general shall be taken <u>as</u> confirmation of any theory (determinism itself being one such) about the world.

We may then ask whether experience of the application of scientific theories (or of other theories about the world) to the world is consistent with the thesis that any

item of experience can in principle be grasped as falling within the context of some potentially recurrent unified sequence pattern. And we have found that to ask such a question is, more practically speaking, asking whether any item in experience can legitimately be regarded as being in principle a value of some variable within some limited system of variables such that it can be calculated given the laws which fully express all relevant changes for that system and given also a specification of values of those variables for some (arbitrarily chosen) initial time.

One thing seems clear immediately: determinism, expressed in this form, does not carry any aura of obviousness, necessity, or "a priority" about it. We have begun by supposing rather vaguely that the world formed a coherent unified orderly whole, an idea initially expressed as the view that each state of the universe is exhaustively determined without remainder, by that which preceded it. But this vague supposition was found to be insufficient alone to entail that within our experiences of that world, finite and limited segments formed peculiarly "connected" sequences. If science has been a paradigm of human "knowledge" for us, or even if some form of "enlightened commonsense" has been our guide, we will have wanted to express an epistemological thesis which would anchor, in the world, the beliefs and expectations which we (currently) have pertaining to the finite and limited segments of

experience which we regard as "connected" sequences. The thesis of determinism at which we have arrived expresses such an epistemological thesis, but in the evolution of the thesis we have abandoned a dimension of the original vague supposition which was, I think, responsible for its initial intuitive plausibility: namely, the postulation of a coherent unity in the world as a whole. In order to provide a philosophical justification for our favourite beliefs and expectations, it is sufficient to anchor them <u>within limits</u> --the limits of the systems of variables which we abstract, and the limits of our abilities to measure real items of experience taken to constitute values of those variables. We need merely regard every item of experience as belonging in principle to <u>some</u> system of variables which would, if it and appropriate covering laws were known, enable us to calculate that item of experience.[10] We do not need to suppose that there is some <u>one</u> all-encompassing "system", in order to successfully rationalize that which we do in the name of science or, in a cruder fashion, in the name of commonsense. We do not need to suppose that between any two finite abstractly conceived systems there necessarily exists some interrelationship which locates those systems in a larger picture.

Now that the thesis of determinism has lost its intuitive plausibility through our loosening up of the concept of a "unity of nature", it can only itself be <u>known</u>

to us in virtue of an empirical fit with experience, particularly those experiences which comprise our knowing activities. But certain problems now arise due to the very conceptual loosening up which we required in order to render the thesis significant with regard to human knowledge. The application of theories of deterministic form in our grasp of the world is now seen to involve at least two non-determinate dimensions. Recall that we first had to require that such theories be relevant only to limited systems of variables or types of variables, all excluded variables to be regarded as irrelevant to the patterns of change which were of interest among those variables selected. The problem introduced by this limitation is that the formulation of an abstract concept of a sufficiently closed system of variables or types of variables requires that we decide which sorts of variables actually do proceed along the lines of relatively independent and self-contained processes of change, that is, that we decide which sorts of variables do constitute such sufficiently closed systems. And this decision must be based on an awareness that for some given set of variables, similar sorts of sequences do regularly recur as such. But in order to reach such a decision, we must, in effect, judge other variables to be irrelevant to the sequences thus apprehended. Yet because we have been accounting for knowledge as a law-governed grasp of items in experience, then it would seem that we

can only so judge legitimately by being able to apply some theory of deterministic form in our very apprehension and grasp of those sequences as sequences.

Application of mechanistically deterministic theories to experience thus requires the abstract isolation of systems of variables; but this requires some prior grasp of relevant variables and significant sequences in experience, and the discrimination of those from irrelevant variables and incidental regularities which shall have no theoretical import in the given context. And this seems to require a prior dependence to begin with on some, albeit crude and rudimentary "theory" of deterministic form. Since we cannot have our theory before we have our theory, we must suppose that we work gradually, by trial and error, or more swiftly by some feat of insight, through a series of ever-more refined notions of abstracted, systematized, and sufficiently closed sets of variables. We may never be warranted in supposing that some given system, which is the object of application of a corresponding set of laws, is the most inclusive and most complete system possible for the purpose. Yet we can still only trust our laws to the extent that we feel our system of variables to be sufficiently complete and sufficiently closed to interference from excluded variables.[11] However it has been claimed[12] that there is no known case of a law which can exhaustively account for the changes occurring in some system of

variables, such that those changes can be regarded as being completely free from dependence on contingencies introduced from outside the system.

The noted circularity of the concepts of theory and system is not necessarily vicious. We can regard the world of our common sense experience as the somewhat arbitrary starting point at which we begin the conscious articulation of unified theory. From this point of departure which is, in a sense, "given" to us, we can allow further perhaps contrived experiences to constitute the occasion for a potentially unending sequence of complementary refinements between our articulated theory and our concept of the abstracted system to which it is thought relevant and applicable. A more perplexing problem consists in the apparently prior "law-ladenness" of that world of common sense experience which is "given" to us as the point of departure for an articulate science. It is the dependence of experience itself upon notions of lawful connection that can be regarded as Kant's basic insight in his attempted resolution, in the Second Analogy of Experience,[13] of Hume's problem of causality. Kant's argument is that in order that appearances can be regarded by us as constituting genuine objective experience (with regard to which alone knowledge is possible), it is presupposed that segments of experience be regarded as being fixed in some determinate order. Such tentative initial

orderings, though Kant is not clear on this, must presuppose the application of notions of regular relationships--crude, rudimentary, pre-scientific "laws". Lawfulness seems to emerge as a presupposition of all experience.

The point has been made before[14] that if all events were random, inexplicable, and unforeseeable happenings, or if there were too many of these, we should be hard put to construct a conception of objective experience of a real and knowable world. This seems beyond question, and it was Kant's insight to have grasped some of the presuppositions inherent in this construction. But it is not altogether clear that for there to be a whole coherent experience for us, every item which is grasped as an item of experience must be grasped as located within some potentially recurrent unified sequence pattern. The unlimited extension of this claim would be impossible of meaningful test. The presuppositions of objective experience are not a matter of conscious awareness. It is therefore necessary to reconstruct them for any particular (apparently) objective experience in accord with relevant logical and epistemological criteria. Hence the legitimacy of such a reconstruction must be decided upon on the basis of prior philosophical commitments as to what has to be involved in the having of objective experience.

Furthermore, the prima facie case itself fails to bear out the unlimited applicability of the requirement of

lawful relationship. And this results from our requirement that notions of lawful relationship be applicable to some system of variables which is intellectually manageable within the limitations of human experience. For if it is true that there is no known law which can exhaustively account for the changes occurring in some system of variables, then we seem doomed to find that the application of our theories is always attended with some degree of theoretically inexplicable variations in the events themselves. In the context of applying the theory at hand, those variations must then and there be regarded as undetermined "items", items of experience for which (in the context at hand) there is no subsumption under notions of lawful relationship. Perhaps the requirement of lawful relationship need not be of unlimited scope in order that a case can be made for its role as a limited presupposition of experience. It seems to be a fact that we can grasp items of experience as belonging to an undetermined experiential fringe, that is, that appearances can constitute objective experience for us without having to be grasped as parts of potentially recurrent unified sequence patterns. It is precisely such prima facie randomness of even mechanical events that constitutes the basis of the attraction of so-called "games of chance". It may be true that given sufficient information, a physicist could accurately predict the results of successive throws of the

dice, or spins of the roulette wheel. But such extensive information is surely not involved at the ordinary level at which such events constitute objective experience for us.

It has been argued[15] that items which are a matter of "chance" relative to some given context may not be a matter of chance in some other context (for example, a future situation in which a more inclusive theory can be applied), so that all items apprehended as matters of chance are really not apprehended as such absolutely. This point is well taken but unfortunately it will not take us very far. The question is whether knowledge, and, more fundamentally, experience itself is at all possible if the items experienced are not grasped <u>as</u> items in lawful patterns. I have granted that no knowledge, and probably no experience, is possible if <u>no</u> item of (potential) experience can be grasped as an item in a lawful pattern. But there is no compelling reason to grant the further more sweeping thesis that no knowledge is possible, nor perhaps any experience either, unless <u>each</u> item of (potential) experience is grasped as an item in a lawful pattern. Indeed the very application of a deterministic theory to experience has been found to require the abstraction of some finite and experientially manageable system of variables to which it is to be thought peculiarly applicable. Thus the application of a theory will involve, relative to the theory and the situation of its application, the non-lawful grasp of

some items which seem nevertheless to constitute genuine objective experience. We may allow, however, that the objective experience of items regarded as non-lawful may only be possible because they are items within broader ranges of experience that, as wholes, can be grasped in terms of some deterministic theory. In any case, the notion of objective experience seems to extend beyond merely that for which a lawful account can be given.

A further problem arises out of the limitations which we have had to recognize in the nature of humanly possible deterministic theories. This problem pertains to the previously noted margins of mathematical divergence of measured values from the precise results calculated by use of the theory. We are forced to regard the application of a deterministic theory as being adequate, and as providing knowledge, just so long as the divergence of measured from theoretically calculated values falls within some tolerable limit. As Frank has argued,[16] the results of measurement are never mathematically precise numbers, but rather mathematical intervals, so that not only the calculated state of the system but the (arbitrary) initial state as well are given by identifying mathematical intervals within which the precise actual quantities fall. To one and the same observed state of a system however, a great many mathematically computed states may correspond. It seems likely that the more remote in time is the calculated state

from the initially given state, the more likely is the prospect that wide margins of difference will appear between the calculated and the measured values of the derived state, due to continued magnification of even initially small differences through each stage of change in the system.[17] Frank suggests that the application of laws to observable states is based upon an assumption that small variations in the initial states defined by sets of intervals of measurement cannot result in large variations in the final state. This seems a likely assumption of refined scientific procedure, though it is less compelling as a general claim of what we find true of experience. And as Frank notes there are many examples of laws on the basis of which calculations are derived which do not bear close approximations to the corresponding observed states of the relevant systems. Again we find that in the very contexts of applying deterministic theories to experience, there are, relative to the situation and the theory then and there applied, some items of experience which are not grasped in terms of any lawful, deterministic patterns.

There is a sense in which anything which is objectively experienced is thereby "known". It is therefore important to demarcate from this sort of knowing that sort with which the thesis of determinism is concerned. The latter must be distinguished as the sort of knowing in which an "account" can be given of that which is known.

It is not a mere perceptual recognition of the characteristics of a thing (experienced); it is the knowing involved in being able to explain. The thesis of determination is thus the thesis that something can only be known, in the sense of being explained or accounted for, if the explanatory account can, on grounds which are independent of the explanatory sequence itself, be recognized as having explanatory relevance. An independent ground which seems immediately plausible is the recognition of the potential recurrence of events such as that which is to be explained together with the same sorts of explanatorily relevant conditions, in the same pattern as that which is exemplified in the given instance. The term 'knowledge' as used in the remainder of this chapter will refer to the sort of knowing in which an account or explanation can be given of that which is known.

In developing a plausible and experientially significant thesis of determinism we have been led through the following stages of reasoning:

1. The notion of a universe wholly coherent, unified, and deterministic throughout had initial plausibility but was empirically meaningless in regard to the limitations of human experience of that universe. The notion therefore required drastic refinement in ways which were not implicit in the original notion itself.
2. It was postulated that determination of the universe as a whole would be recognizable to human beings if, within the narrow ranges of human experience, there were finite and recognizably determined sequences of items of experience.

3. The recognition that items of experience constituted unified determined sequences, and not merely incidental arrangements, required, at least, that those sequences be grasped in terms of abstract patterns which were potentially recurrent as such.
4. The practical recurrence of abstract patterns seems to require corresponding abstract notions of isolated sets of variables for which the patterns could be regarded as expressing systematic lawful interrelationships.
5. The mathematically precise expectations entailed by a theory for an abstracted system of variables must be attributed a successful empirical application even within a range of divergent measured values.

But it appears that, in achieving knowledge of experience through the justified application of some deterministic theory to some finite range of unabstracted experience, there are simply going to be items of experience within that range which are not grasped, within that context, as parts of finite recognizably determined sequences--yet still constitute objective experience. Knowledge and objective experience both seem quite possible for human beings, with regard to ranges of unabstracted experience as wholes, so long as a deterministic "grid" can be legitimately regarded as applicable to those ranges, that is, discerned within those ranges of experience. It does not seem to be additionally required that the discerned "grid" incorporate each and every item of experience.

At this point we might postulate the thesis of determinism as a <u>regulative ideal</u> of the scientific

enterprise, a directive which enjoins a certain approach to the study of the world thought to be the most fruitful in terms of the standards of success definitive of that enterprise.[18] This is to say that we would abandon any formulation of determinism according to which every item in experience can be grasped in terms of lawful pattern. As a regulative ideal, determinism would instead consist in the thesis that we (simply) should seek to grasp as much of experience as is possible in terms of lawful pattern. Unfortunately, if determinism is a thesis which at best can only express what we should seek to do, in any concerted effort to achieve knowledge of the world, then we cannot get beyond this thesis itself to discover a justification for following this particular directive rather than any other. If determinism only enjoins a method for studying the world without presupposing any commitments as to the way the world is and, hence, what can in fact be achieved by studying the world in that way, then determinism remains a thoroughly arbitrary thesis. It does not even entitle us to say that we are more likely to achieve knowledge of the world by seeking lawful patterns among the phenomena, than if we study the world in some other manner.[19] For such an appeal would involve the tacit assumption that, more often than not, items of experience actually exhibited lawful patterns--an assumption which, as we found, cannot be tested beyond the prima facie case and one which,

moreover, is not clearly borne out by that case.

In order that there can be knowledge in a full but practically relevant sense, we need only require that items within ranges of experience be grasped as parts of potentially recurrent unified sequence patterns, and this grasp seems to require the abstract isolation of systems of variables of which the given items can be regarded as values. Consequently, it is required that for any item of experience to be known, there must be other items of experience which additionally comprise the concept of the relevant system of variables which is brought to bear in understanding the case at hand. Thus to grasp any item of experience as instantiating a lawful pattern, one must, at the same time, be aware of lawful pattern as pertaining to a plurality of other items in experience. But this systematic nature of knowledge, to reiterate and insist upon the point, does not lead us anywhere near a presupposition that all items of experience are, or need be regarded as, parts of lawful sequences.[20]

David Bohm has articulated a "point of view that goes beyond mechanism" and which seems adequate to express the guiding ideals of the scientific enterprise without entailing a certain narrowness of focus and expectation that he regards as characterizing the mechanistic viewpoint.[21] According to this point of view, the general structure that has thus far actually been found in the laws

of nature has the following essential feature:

> Any given set of qualities and properties of matter and categories of laws that are expressed in terms of these qualities and properties is in general applicable only within limited contexts, over limited ranges of conditions and to limited degrees of approximation, these limits being subject to better and better determination with the aid of further scientific research.[22]

Beyond these limitations on the validity of any given theory, the possibility always remains open that there exists an unlimited variety of different kinds of things (properties, qualities, entities, systems, levels) to which only new kinds of laws of nature will be applicable. Yet this notion of, in Bohm's words, "the qualitative infinity of nature" does not involve abandonment of any insightful consequences of the mechanistic philosophy. In particular, there is no need to abandon the basic mechanistic insight that broad ranges of prima facie independent phenomena can be grasped as one unified domain of phenomena in terms of some significantly finite and small set of general qualities, properties, laws, principles, etc. What we gain from Bohm's notion is a measure of realistic caution in the attempt to apply existing theories to the search for accurate predictions in new domains, in new contexts, and to new degrees of approximation. To quote Bohm once again:

> . . . we are led to a concept of the nature of things which is in complete accord with the most basic and essential characteristic of the scientific method; i.e. the requirement of continual probing, criticizing and testing of every feature of every theory, no matter how fundamental that theory may seem to be.[23]

Following Bohm, we may say that determinism can no longer be regarded as an affirmation of the universal lawfulness of all items in experience as a necessary presupposition of the possibility of human knowledge. Perhaps we can turn our previous formulations of this thesis on their heads by expressing determinism in terms of what is inexplicable and unknowable. Thus determinism can be expressed as the thesis that nothing which occurs in a truly chance fashion, i.e., lawlessly or randomly, could be known or grasped as such. If knowledge is only possible by means of apprehension in terms of lawful pattern (here regarded as encompassing probabilistic sequences), then where there is no lawful pattern, there can be no knowledge.

A plausible argument for such a view, indeed one which could be applied to every generalization, might proceed as follows:[24] Thoroughly random items in experience would be items for which, by definition, no known, or knowable, law could enable us to calculate their values, no matter what the specified initial values. However the "characteristic" of failing to be lawfully patterned, that is, failing to be lawfully related to any other items in

experience is not one which could ever be positively ascertained, for all that we could intellectually manage in regard to such an item of experience is a grasp of our failure to discern any lawful pattern. But we can fail to discern a pattern in experience even if there is a pattern to be discerned, and even if we currently possess the intellectual resources sufficient to enable its discernment. Consequently our grasp of our failure to discern the place of some item in a lawful pattern is not in itself sufficient to imply the genuine randomness of the item in question. We can never be justified in (positively) attributing randomness to any item in experience. The most that can be said is that for various items in experience in various contexts, there may well be a simple absence of attribution of positive relationships with any other things.

FOOTNOTES

[1]Noted also by Philipp Frank, Philosophy of Science (Englewood Cliffs, 1957), p. 282; and Arturo Rosenblueth, Mind and Brain (Cambridge, 1970), p. 275.

[2]Gustav Bergmann alludes to a sense of "deterministic" which involves extending to the "whole world" the notion of comprehensive lawfulness. Such a law might be taken to refer to some one, and only one, "gigantic process" encompassing all variables. Bergmann is non-committal regarding this construal of "deterministic". Cf., "Explanation, Prediction, and Imperfect Knowledge", in Mary Brodbeck, ed., Readings in the Philosophy of the Social Sciences (New York, 1968), p. 418.

[3]We are told this, for example, by Bertrand Russell, "On the Notion of a Cause" in Mysticism and Logic (Garden City, N.Y., 1917), pp. 196-197; by Moritz Schlick, "Causality in Everyday Life and Recent Science", in Herbert Feigl and Wilfrid Sellars, eds., Readings in Philosophical Analysis (New York, 1949), pp. 527-528; and by Carl Hempel, "Some Reflections on 'The Case for Determinism'", in Sidney Hook, ed., Determinism and Freedom in the Age of Modern Science (New York, 1958), p. 159.

[4]In "Every Event Has a Cause", in Antony Flew, ed., Logic and Language, second series (Oxford, 1961), p. 108.

[5]Cf. Bergmann, op. cit., p. 416: "From our paradigm [celestial mechanics] and the thermodynamic example we know that there are temporal sequences of events believed to be in fact processes for reasons as excellent as one can ever hope to have within the inescapable limits of measurement and inductive 'uncertainty'. Of course this is a matter of fact, not of logical or any other 'necessity'". (Bracketed insertion is the editor's, not my own.)

[6]For a similar point, cf., Frank, op. cit., pp. 269-274, 282-284; Russell, op. cit., pp. 189-192; William Kneale, Probability and Induction (Oxford, 1949), pp. 64-65; Ernest Nagel, The Structure of Science (London, 1961), pp. 280-281.

[7]Although Schlick, for example, has avoided the notion of predictability altogether in favour of the notion of calculation. Cf. op. cit., p. 525.

[8]The point that determinism does not entail predictability is made by, for example, Sidney Hook, "Necessity, Indeterminism, and Sentimentalism", in Hook, ed., op. cit., p. 167; Paul Weiss, "Common Sense and Beyond", in Hook, ed., op. cit., p. 222; Alvin Goldman, A Theory of Human Action (Englewood Cliffs, 1970), pp. 178-180; Adolph Grünbaum, "Free Will and Laws of Human Behaviour", American Philosophical Quarterly, vol. 8 (1971), p. 300.

[9]The claim is put forward by, for example, Frank, op. cit., pp. 272-274; Bergmann, op. cit., p. 416.

[10]For a similar point, cf. Frank, op. cit., pp. 282-285; Nagel, op. cit., pp. 322-323.

[11]As Stephen Korner notes, the application of a theory to an empirical situation presupposes--though doesn't guarantee--that features of the empirical situation which have no counterpart in the theory are not only neglected, but are negligible--not only theoretically irrelevant, but, in that context, ineffective. Experience and Theory (London, 1966), p. 221.

[12]For example by David Bohm, Causality and Chance in Modern Physics (Philadelphia, 1957), pp. 2-3, 20-21, 61.

[13]Critique of Pure Reason, Norman Kemp Smith, trans., Second Analogy: A 189-211, B 232-256. Frank interprets the point more specifically with regard to experimentation. He regards Kant as having argued that we take a "causality principle" for granted in our very regard for an experiment as revealing to us certain properties of a body; the experimental results must be regarded as "effects" attributable

to properties of the body regarded as "causal". Our very definition of states of a system, such that universal laws can be set up for sequences among them, presupposes the discernment of regularity in the experimental discrimination of those states. Frank, op. cit., pp. 282, 286-288. Unfortunately this insight has not been sufficiently recognized. Cf. W. A. Suchting, "Kant's Second Analogy of Experience", in Lewis W. Beck, ed., Kant Studies Today (La Salle, 1969), pp. 322-340, esp. p. 335; and the otherwise insightful Peter F. Strawson, The Bounds of Sense (London, 1966), pp. 146-149. Gerd Buchdahl recognizes that for Kant the concept of causality is a presupposition for the possibility of experience in general, but feels that at this level its function is merely to bring it about that the elementary sense constituents of objects of experience are thought of as members of some (as yet undetermined) necessitarian causal sequence. Cf. Buchdahl, "The Kantian 'Dynamic of Reason' With Special Reference to the Place of Causality in Kant's System", in Beck, ed., op. cit., pp. 341-374, esp. pp. 355-360; and Buchdahl, "Causality, Causal Laws and Scientific Theory in the Philosophy of Kant", British Journal for the Philosophy of Science, vol. XVI, pp. 187-208, esp. pp. 193-197. An even more agreeable interpretation is to be found in Lewis White Beck, "Once More Unto the Breach: Kant's Answer to Hume, Again", Ratio, vol. IX (1967), pp. 33-37, although here again Kant is interpreted as requiring only that those representations, to be taken as representations of events, be those whose order is thought to be fixed by (some) other temporally fixed events which need not be picked out specifically. But without determinately fixing given representations through application of some concept of a specific lawful relation, we cannot fix the order of those representations so as to regard them as constituting an objective realm of experience. More agreeable yet is the interpretation offered by Jonathan Bennett, Kant's Analytic (Cambridge University, 1966), pp. 219-229, though Bennett makes the mistake of separating two strands of the argument which must be allowed to remain intact for mutual support. For a full discussion of the Second Analogy, cf. my "Kant's Second Analogy: The Metaphysics and the Epistemology of Causality", unpublished.

[14] Cf. Warnock, op. cit., pp. 107-108.

[15] Cf. Nagel, op. cit., pp. 329-330.

[16] Op. cit., pp. 275-277.

[17]For a similar point cf. Robert Ackermann, The Philosophy of Science (New York, 1970), p. 104.

[18]Cf. Nagel, op. cit., pp. 317-324.

[19]As, for example, Hook says, even while acknowledging that, expressed in this way, determinism is nothing but a postulate or hope; op. cit., p. 168: "If we act on the assumption that a system is determined, then it is more likely that we shall discover laws and make successful predictions about the future than if we assume that chance reigns". C. S. Peirce has noted that if mechanistic determinism is regarded as a "presupposition" or postulate of scientific reasoning, then this is to say nothing more than that it is hoped to be true. Cf. "The Doctrine of Necessity Examined", in Vincent Tomas, ed., Charles S. Peirce: Essays in the Philosophy of Science (Indianapolis, 1957), pp. 172-173.

[20]Peirce notes that observations generally adduced in favour of mechanical causation simply prove that there is an element of regularity in nature, but have no bearing on the question of whether such regularity is exact and universal or not. Cf. Peirce, op. cit., p. 179. Warnock has noted that while it is true that if there were too many random, inexplicable, and unforeseeable happenings we would be in practical and linguistic difficulties, nevertheless this does not show that we cannot tolerate any disorder or chaos at all; op. cit., p. 108.

[21]Op. cit., esp. pp. 132-136, but also pp. 137-164.

[22]Op. cit., p. 133.

[23]Op. cit., p. 136.

[24]Cf. some brief remarks by Nagel, op. cit., pp. 331-335.

III

CAUSALITY

The thesis of determinism, as it is worked out in the preceding chapter begins with the postulation of a certain necessary structuring of our regard for the world of our experience in order that that regard constitute knowledge of that world. The necessary structuring is one of potentially recurrent lawful pattern. Viewed in this way, the thesis of determinism is an essentially epistemological thesis. But the deterministic thesis, in the essentially epistemological formulation which it has been given, places no restrictions upon the nature of the patterns which human beings are required to discern in the world around them in order to be able to understand that world. In particular, the deterministic thesis does not imply that the only patterns to be discerned in the world are causal patterns,[1] patterns according to which that which we infer as a logical consequent (in accord with the law expressing the pattern) of the remaining parts of the pattern taken as its antecedent determining conditions is also regarded as a causal consequence of those antecedent conditions.

In essence, the deterministic thesis, aside from

expressing a hope as to what human intelligence and endeavor can accomplish in the future, is little more than a thesis about the logical nature required of understanding in order that it constitute genuine knowledge. It is required that explanations, as distinct from mere descriptions of events, must be implicitly generalizable. An explanation for some phenomenon should reveal that phenomenon to be part of a sequence of phenomena which is itself an instance of a potentially recurrent unified sequence pattern. To grasp some phenomenon in terms of such a lawful pattern is, in effect, to apprehend the phenomenon within an intellectual context in which it is <u>necessarily</u> regarded as occurring with a given degree of probability. Note: I do not say that it is regarded as <u>occurring necessarily</u>. For the epistemological thesis of determinism requires nothing more than the following neo-Humean account of necessity.[2]

The experiential sequences which exhibit the patterns regarded as explanatory are not thought to exhibit these patterns by virtue of any necessity inherent in the sequences themselves but only as a matter of coincident constancy which we find to obtain through experience of other relevantly similar sequences which do instantiate the relatively same abstract patterns. The apparent "necessity" with which we regard the explained phenomenon as occurring is, in actuality, the necessity of <u>that which we are to do</u> in regard to this item in a particular intellectual context,

namely, infer its occurrence in accord with the law formulating the relevant explanatory pattern once the occurrence of the antecedent conditions of that pattern is given. The aura of necessity associated with the sequences which we regard as lawfully patterned is thereby construed as nothing more than the secured nature conferred upon those lawful patterns in our thinking by their systematic interconnection in a body of scientific theory which we accept as a whole and upon which we are prepared to rely in formulating plans and expectations for the future. This systematic theoretical interconnection sufficiently distinguishes laws of nature from other regularities which we find to obtain in experience but which we do not regard as being "necessary", as being explanatory of the phenomena which comprise them. Furthermore it is our preparedness for a future which conforms to the systematic theories which we adopt, a preparedness which is part of what constitutes the adoption of a theory, that underlies our readiness to accept counterfactual and subjunctive statements based upon the laws comprising the adopted theory even while we refuse to accept such statements with regard to merely observed, and therefore (as far as we are concerned) thoroughly coincident regularities.

In accord with this view, there is no particular scientific importance to be attached to the ordinary concept of causality. This concept must be regarded as

having essentially practical or common sense usefulness only. Salient features of the ordinary concept of causality are as follows:

1. Some specific event or entity is selected from among surrounding conditions as constituting "the cause" of a selected event or entity which is its "effect".
2. The sequence is probably regarded as being of an invariably uniform sort, the effect regularly following upon its cause.
3. There is spatial and temporal continuity between the cause and the effect.
4. The relationship between the cause and the effect is asymmetrical.

The uselessness of the ordinary concept of cause for any sophisticated study of nature depends on a number of difficulties regarding these constitutive features.[3] To begin with, there does not seem to be any invariably uniform sort of sequence consisting of only a single event or entity taken as cause and a single event or entity taken as effect. Sequences regarded as having this nature are in fact being regarded within the context of an assumption that certain background conditions persist throughout the occurrence of the sequence. It is thus only in conjunction with such background conditions that any single event or entity can give rise to an effect with anything like invariable uniformity.

A reconsideration of point (2) thus leads to a reconsideration of point (1). If a "cause" is only causally efficacious in conjunction with a set of background

conditions, then such conditions become as necessary to the production of the effect as is the condition isolated as "the cause". All factors being thus equally necessary, and none by itself sufficient to produce the effect, it becomes a matter for speculation as to why the isolated factor should ever have been singled out in this way and why, at any rate, it should continue to be so regarded. The only accounts that plausibly explicate our discrimination in these matters reveal this discrimination to be a thoroughly contingent anthropocentric and scientifically irrelevant matter. Conditions are thought of as background conditions because from our point of view in the normal situation of our lives, they remain largely or entirely unchanged, at least in their gross observable characteristics. We take their obtainment for granted to the point where we fail to explicitly acknowledge their necessary contribution to the entire state of affairs in which alone some causal sequence is possible. The factor then explicitly regarded as "the cause" is one which has some peculiar or interesting feature from our point of view <u>within</u> the situation in which the causal sequence occurs. For example, it may be the factor which stands out against other factors as figure to ground in our perceptual organization of the situation, perhaps because of apparent motion in relation to other factors. Or it may be the factor which is most readily available for our direct manipulation in the sorts of

situations which we encounter practically.

A reconsideration of point (1) leads to a reconsideration of point (4), which pertained to the asymmetry between cause and effect. There seems to be an intuitive plausibility to this notion but its explicit clarification is not easily had. It might be thought that this asymmetry is of a logical nature, that is, that it pertains to our inference habits in regard to a given cause-effect sequence. Thus given a specified cause, the effect would be uniquely "determined", in the sense that it would be uniquely determined as to what was to be inferred as being the effect of the specified cause; but given a specified effect, the cause would not be uniquely determined. The causal relationship would thus be conceived as a logical "many-one" relationship in that for any cause a unique consequence could be derived, but for any effect, no unique cause could be selected from among an inferred plurality of possible causes. Thus, if a given set of causal conditions, say, $C_1 \ldots C_n$, produce a given effect, say, E, then we shall be able to infer the effect E given a specification of the causal conditions. But we shall not be able to infer the cause given a specification of the effect because the same effect could also occur as the effect of an entirely different set of causal conditions, say, $C'_1 \ldots C'_n$.

But the replacement of the notion of cause as a single entity or event by the notion of cause as a set of

(determining) conditions indirectly undermines the plausibility of this explication of causal asymmetry. The concept of a cause has become a systematic concept, one defined for a set of causal conditions. For such a set of factors, the genuine "effect" of their conjoint causality is not simply a single event or entity but a set of "effected conditions" comprising the total resultant across the entire range of variables of which the initial set of causal conditions are values. Thus, suppose that the effect is specified as a set of effected conditions, $E_1 \ldots E_m$. Then there is no more reason to suppose that the same effect could occur as the effect of several entirely different sets of causal conditions than there is to suppose that the same set of causal conditions could give rise to a variety of effected conditions. Under this interpretation, not only does the cause uniquely determine its effect, but the (fully specified) effect uniquely determines its cause.

It may be thought[4] that there is still a logical asymmetry in the causal relation, even if the effect is conceived as a set of effected conditions uniquely determining its cause. Suppose that a given set of causal conditions, $C_1 \ldots C_n$, causes a given set of effected conditions, $E_1 \ldots E_m$. Then the following inference holds true:

$$C_1 \ldots C_n \supset E_1 \ldots E_m$$

as well as any of the following:

$$C_1 \ldots C_n \supset E_1$$
$$C_1 \ldots C_n \supset E_2$$
$$\ldots$$
$$C_1 \ldots C_n \supset E_m.$$

But the following sorts of inferences do not hold true of the causal relationship:

$$E_1 \supset C_1 \ldots C_n.$$

Superficially, it appears that a logical asymmetry indeed obtains in regard to causal sequences.

But if the effect is conceived as a set of effected conditions, and if an effect conceived in this way does indeed uniquely determine its cause, then the following inference also holds true of the causal relationship:

$$E_1 \ldots E_m \supset C_1 \ldots C_h,$$

and, on the basis of this, the following sorts of inferences also obtain:

$$E_1 \ldots E_m \supset C_1$$

while not any of the following

$$C_1 \supset E_1 \ldots E_m.$$

The original "asymmetry" of permitted inferences is, we might say, itself symmetrical as between cause and effect, so long as these are specified in identical form.

It would seem plausible to think that if there is an ontological difference between cause and effect, something which grounds the apparent asymmetry of causal sequences, then this matter of ontology would be reflected in the logical form of descriptions of the causal sequences. But so long as the restrictions as suggested here, on how we are to formalize the cause and the effect, are adhered to, then there does not appear to be any asymmetry at the logical level. At most, we can say only that, due apparently to ignorance on our parts, we do not ordinarily conceptualize effects as sets of (effected) conditions, so that, in fact, we deal with single effect conditions which may then be implied (as part of various total sets of effected conditions) by a variety of different sorts of sets of causal conditions. Perhaps, then, there is an asymmetry in our knowledge of diverse sorts of causal sequences, taken as an unsystematic conglomerate. However, even here, it may also be said that in fact we often do not conceptualize causes as sets of (causal) conditions. Rather, we deal with single causal conditions which may then be regarded as capable of implying a variety of different sorts of effects. True, we acknowledge that the precise sort of effect produced depends upon the nature of the background conditions which (also) obtain, in conjunction with the selected causal condition. But my point has been that were we to follow through consistently on the implications

metaphysical interpretation to be given some mathematical calculus. For just as a given set of observations do not uniquely select a single mathematical calculus as formulating the relationships among those observations, so too a given mathematical calculus does not uniquely select a single metaphysical interpretation of itself. It has therefore been thought by many[13] that in light of such considerations as those just noted and because, in any case, metaphysical interpretations exceed the intellectual "structuring" requisite for the possibility of scientific explanation, prediction, and derivative technological control, that they are therefore, on all intellectual levels, irrelevant and dispensable to a grasp of scientific theories. Even if the previous considerations are so problematic as they seem, this conclusion seems excessively restrictive. As Frank suggests, the untestable nature of so-called "metaphysical" views may only be relative to certain formulations of them.[14] It may therefore be relative to the conceptual equipment of a certain period of time. Surely it is possible that we can progress not only in the formulation of new mathematical calculi with which to calculate ever increasing aspects of the world, but as well in the "operationalizing" of evermore previously untestable views about that world.

Metaphysical viewpoints may serve a further practical function as suggested by Frank. They may provide

of this approach, then we would have to say that a single effected condition implied a precise and unique sort of cause, depending on which effected background conditions (also) obtained, in conjunction with the selected effect condition. Any logical asymmetry between cause and effect is thus only apparent and derives from an unduly narrow specification of effect in relation to a more broadly specified cause.

As a second approximation to what seems intuitively at issue here, it might be thought that the asymmetry between cause and effect--both conceived in their entirety--is of a temporal nature, that is, that the causal conditions must precede the effectual conditions in time. Thus it is acknowledged that given a complete specification of the cause, a unique inference of its effect is determined, and given a complete specification of the effect, a unique inference of its cause is determined, the logical relationship being thus perfectly symmetrical. But it is nevertheless thought that in the course of the occurrence of real sequences which instantiate any such causal pattern, one set of conditions will always occur first in time. And this set will therefore properly be called the "cause" of the other set.

The difficulty about this suggestion is that the logical relationship, which underpins our concept of the relationship of determination, has been "used up" in that

as occurring "for no reason". Their inexplicability in this sense would, I believe, prevent us even from regarding such actions as actions for which the persons could be held personally responsible. Where those actions were harmful or detrimental, we should regard them as evidences of insanity or abnormality, requiring treatment rather than punishment or condemnation. Yet, as I shall argue in a subsequent chapter, the interest in notions of moral responsibility, and the concept of man as a moral agent, is one of the strongest factors behind the postulation of free will. Thus if free will were conceived as requiring undetermined, that is, unlawful or random, human action as its manifestation, then the postulation of free will would be simply unacceptable on moral (as well as epistemological) grounds. The only acceptable conclusion is that pattern and determination are essential even to our grasp of such human behaviour as we might wish to regard as freely willed.

In order to understand how the postulate of free will is not inconsistent with determinism but rather with a certain metaphysical interpretation of determinism in terms of causality, it will be useful to examine the basis of this metaphysical interpretation. Recall that there seemed to be nothing about the epistemological considerations underlying determinism which led to any conclusions one way or the other about the <u>nature</u> of the regularities in the world, whether they are causally necessitated, or

possible knowledge for us.

The following account of the asymmetry of causal relations might seem plausible. (I shall subsequently raise some criticisms of it, pp. 103-106 .) It might seem that our everyday regard for temporally antecedent conditions as explanatory factors consists in practical matters having to do with our ability to control and manipulate events, not only in the circumstances of everyday life but even in the laboratory. The condition which is regarded as the cause is grasped as being directly manipulable in the situation, as a means for producing or bringing about some other condition, therefore regarded as the effect. Actual manipulability by human beings may be beyond the technology available at a given time, as, for instance, in the case of changes among geological or astronomical phenomena. But we may still, in such cases, organize our grasp of the phenomena in terms which reflect an action-orientation that appears potentially available to us among the phenomena themselves should we someday develop the technological capacity to avail ourselves of it. What we are grasping is the availability of one condition, against suitable background conditions, to be directly changed as a means of altering some further condition. And our practical activity with regard to conditions is quite frequently asymmetrical.

Thus our <u>practical</u> grasp of a lawful pattern, that is, our apprehension of it in its relevance and application

to our experiences, involves a fitting of it to the circumstances of our environment as they manifest themselves during the course of our interaction with this environment. A generalized statement of a lawful regularity among otherwise neutral conditions becomes, from this practical point of view, a statement involving, as Wilfrid Sellars puts it, an objective distinction between standing conditions, a doing, and a result or consequence.[8] Sellars characterizes this application-relevant regard for causal generalizations by saying that ". . . one explains a particular matter of empirical fact by 'showing how it comes about' and not, simply, by subsuming it under the 'consequent' of a general hypothetical, the 'antecedent' of which it is known to satisfy. . . ."[9] How the sequence is regarded as "coming about", in turn, depends on what manipulations we are able to perform on given conditions and entities, and what events can be achieved, in the given contexts, only as a consequence of these manipulations. The factors which thus underlie our search for causes appear to be non-scientific pragmatic factors having to do with the incidental constraints upon our abilities to accomplish intentional actions in the world which we attempt to know.

We may thus draw a distinction between the temporal asymmetry which we introduce into our interpretation of and application of deterministic theories and the theories

themselves, which license calculations and derivations that are logically symmetrical and to which the temporal direction of actual events covered by the theories is not logically relevant. The causal relevance of such theories appears to arise in the course of their practical application and is not a matter of the logic of the theories as such. Furthermore causality seems to be a metaphysical interpretation of the logically based relationship of determination which alone is required of those theories in order that they constitute knowledge for us. This important point deserves elaboration.

It is familiar to distinguish, with regard to any sophisticated scientific theory, on the one hand, the pure mathematical calculus of the theory, together with the operational definitions that correlate the calculus with empirical situations from, on the other hand, the metaphysical interpretations that may be put upon this calculus in attempts to render the theory more plausible, more "intelligible".[10] The mathematical calculus enables prediction of the values of the system of theoretically relevant variables at any time, given a specification of the values of these variables for some arbitrarily chosen time. By means of operational definitions and bridge laws establishing correlations between the values of the variables, as described in terms of the theory, and actual phenomena, the mathematical calculus can be applied to

experience in order to "explain" and "predict" further phenomena. The given phenomena, on the basis of which the explanation or the prediction is to proceed, can be regarded as being antecedently given in a logical sense. That is, it is because the occurrence of these antecedent conditions is initially posited that there can be a determination of our minds to infer the occurrence of the explanandum, not necessarily at a time subsequent to that of the occurrence of the antecedent conditions.

However it may seem insufficient that we be equipped merely with an uninterpreted mathematical calculus which enables us intellectually to "get" from a state of affairs taken as logically given to another state of affairs taken as determined by it.[11] In addition, an interpretation may be wanted of the sort of calculations enabled by a given theory, an interpretation which can be regarded as revealing the basis for attributing explanatory power to the theory. Such interpretations can be called "metaphysical" which indicates, among other things, that as formulated at a given time, they are not themselves experimentally testable. Their point may generally said to be that of showing the mathematical calculi of scientific theories to be either intrinsically plausible or to be themselves derivable from intrinsically intelligible and plausible general principles, such as the principle of sufficient reason.

It seems reasonable to think, as Frank argues, that the principles regarded as intelligible and plausible are usually those which reflect some analogy with experiences which are very familiar to us in ordinary practical contexts.[12] However, if this is the case, then certain problems arise in the course of devising such metaphysical accounts for uninterpreted calculi. If the account is based upon the use of some common familiar experience as a "model" for the uninterpreted calculus, one which is "intuitively plausible" just because it is so common and familiar, then it would seem that what each person will be able to regard as plausible and intelligible will be determined not by anything inherently plausible and intelligible but rather simply by that person's fortuitous accumulation of experiences in life. Furthermore, since the terms in which common experience is conceptualized at any one time reflect to some, perhaps belated and confused, degree the advances of science, then the familiar experiences which constitute potential models for metaphysical interpretations of scientific theories are biased to begin with in favour of certain theoretical commitments which shall therefore seem more intelligible and plausible because they already contribute to the terms in which common familiar experiences are already being conceived.

Most problematic of all, perhaps, is the untestability of the choice of analogy upon which to base the

metaphysical interpretation to be given some mathematical calculus. For just as a given set of observations do not uniquely select a single mathematical calculus as formulating the relationships among those observations, so too a given mathematical calculus does not uniquely select a single metaphysical interpretation of itself. It has therefore been thought by many[13] that in light of such considerations as those just noted and because, in any case, metaphysical interpretations exceed the intellectual "structuring" requisite for the possibility of scientific explanation, prediction, and derivative technological control, that they are therefore, on all intellectual levels, irrelevant and dispensable to a grasp of scientific theories. Even if the previous considerations are so problematic as they seem, this conclusion seems excessively restrictive. As Frank suggests, the untestable nature of so-called "metaphysical" views may only be relative to certain formulations of them.[14] It may therefore be relative to the conceptual equipment of a certain period of time. Surely it is possible that we can progress not only in the formulation of new mathematical calculi with which to calculate ever increasing aspects of the world, but as well in the "operationalizing" of evermore previously untestable views about that world.

Metaphysical viewpoints may serve a further practical function as suggested by Frank. They may provide

models for human beings to use in directing the course of their lives. Thus Frank claims that

> We consider that human society is, in a way, a picture of the universe, that we act in a natural way if we act according to the laws of the universe. Man has the belief that as he formulates the general structure of the universe, men in general will imitate this structure in a certain way in their lives.[15]

Thus metaphysical commitments can constitute the basis for recommendations of a better way of life. If this suggestion seems too abstract, one need only consider how the thesis of causal determinism influences our moral thinking and our concept of man as a moral agent.

II

Against the background of the foregoing remarks, I am finally able to present the main thesis of the present chapter. This is the thesis that the philosophical thesis of determinism, which is formulated as an essentially epistemological thesis, is irrelevant to the question of whether or not there exists free will. I do not hereby subscribe to a version of "soft" determinism, the view that within the context of (presumably universal) causal determination, the existence of free will is manifested in those cases in which the causal conditions of an action include an act of choice or volition. What I am maintaining is more radical than this. The notion of causation can be

regarded as a metaphysical interpretation of the deterministic relation. And just as a metaphysical interpretation of a mathematical calculus may be regarded as something over and above, and not uniquely determined by, that calculus itself, so too a metaphysics of causality, as an interpretation of that structuring of the world which constitutes its lawfulness, or determination, may be regarded as something over and above, and not uniquely determined by, the (epistemological) thesis of determinism itself. It is thus possible to provide alternative metaphysical interpretations of the deterministic relation; in particular it is possible to postulate a metaphysical interpretation for, at least, some of the regularities in human behaviour which shall more closely correspond to the intuitive notion which we have of freely willed action. Thus the view presented here is that freely willed action is not to be conceived of as a certain sort of caused event (as soft determinism would entail), but rather as a sort of event which precisely is uncaused--even though it is not undetermined.

Prima facie, the intuitive notion of freely willed action is a notion of action that occurs contra-causally. It is frequently thought that free will can only be postulated at the expense of denying that deterministic theories can be formulated with respect to (freely willed) human behaviour. Thus the free will thesis is frequently

regarded as an implicitly indeterministic view. However I think that such a ramification is forced upon advocates of the free will hypothesis by a misleading expression of the deterministic thesis. The only philosophical thesis of determinism which matters in establishing the possibility of explanation, prediction, control, and the entire scientific enterprise, is the epistemological thesis outlined above, and formulated in terms of the implicit generalizability of our inference patterns and the corresponding lawfulness of the world. It is my basic contention that the intuitive contra-causal notion of free will can be made sense of quite readily within the context of the deterministic thesis. It is the frequent and misleading expression of determinism as a thesis of specifically causal determinism that limits, from the outset, the conceptual tools available to "free will" advocates and forces upon them indeterministic commitments which are by no means implicit in the initial intuitive notion. By equating determinism with causal, or mechanistic, determinism, the "free will" advocate can only deny that human behaviour is exhaustively caused by antecedent conditions by denying that it is determined, i.e. regular or lawful.

But such a denial of the lawfulness of human behaviour is untenable on the grounds of the same epistemological considerations which made determinism a compelling thesis generally. In general, the thesis amounted to the

claim that to the extent that phenomena were undetermined and fit into no discernible pattern with phenomena surrounding their occurrence, to that extent, they would be incomprehensible and inexplicable for us. These considerations apply to all phenomena. To the extent that human actions fit into no discernible patterns of relationship with the conditions surrounding their occurrence, to that extent they would be incomprehensible and inexplicable for us. Yet comprehension and understanding of human behaviour confronts us as a matter of plain fact, not only in scientific contexts but in the course of our everyday lives. We do explain, predict and control human behaviour to a significant degree. These facts simply cannot be disregarded by any account of human behaviour which is to constitute an adequate account. There would seem to have to be some basis for our very frequent successful expectations with regard to the actions of people around us. And if the possibility of such comprehension presupposes lawful patterning of the comprehended phenomena, then there would seem to have to be regularities in the sequences of phenomena which comprise human behaviour, even such behaviour as is freely willed (if there is any).

Furthermore, if human actions were undetermined, having no regular relationship to the beliefs, wants, or character of the persons who performed them, we should regard them as being random, chance events, and therefore

as occurring "for no reason". Their inexplicability in this sense would, I believe, prevent us even from regarding such actions as actions for which the persons could be held personally responsible. Where those actions were harmful or detrimental, we should regard them as evidences of insanity or abnormality, requiring treatment rather than punishment or condemnation. Yet, as I shall argue in a subsequent chapter, the interest in notions of moral responsibility, and the concept of man as a moral agent, is one of the strongest factors behind the postulation of free will. Thus if free will were conceived as requiring undetermined, that is unlawful or random, human action as its manifestation, then the postulation of free will would be simply unacceptable on moral (as well as epistemological) grounds. The only acceptable conclusion is that pattern and determination are essential even to our grasp of such human behaviour as we might wish to regard as freely willed.

In order to understand how the postulate of free will is not inconsistent with determinism but rather with a certain metaphysical interpretation of determinism in terms of causality, it will be useful to examine the basis of this metaphysical interpretation. Recall that there seemed to be nothing about the epistemological considerations underlying determinism which led to any conclusions one way or the other about the <u>nature</u> of the regularities in the world, whether they are causally necessitated, or

whether they involve anything which corresponds to any of the terms which, for example, Hume listed as synonymous with 'causality', namely: ". . . efficacy, agency, power, force, energy, necessity, connexion, and productive quality . . . ."[16] The necessity which we may attribute to a lawfully patterned sequence is, in effect, the necessity with which we must draw a conclusion pertaining to the sequence in accord with the relevant law.

Up to this point, the burden of the present chapter has been that of distinguishing anthropomorphic metaphysical interpretations accompanying scientific theories from the bare epistemological structuring of thought and world which alone seems required in order that we be able to accomplish the "goals" of science: namely, explanation, prediction, and practical control. It has not been my intent to condemn or criticize anthropomorphic metaphysical interpretations. On the contrary, as indicated at the beginning of this section, I shall recommend a particular kind of metaphysical interpretation for the understanding of certain regularities in human behaviour; a particular kind of metaphysical interpretation which is to rely upon a distinctly non-causal model. Consequently, in order to more fully grasp the implications of this recommendation, it is necessary to consider what sort of metaphysical model is being here rejected as a model for understanding those behavioural sequences.

Exactly what sort of interpretation of the relationship of determination, or regularity, does a causal metaphysics provide? It has been noted that metaphysical interpretations of the formal calculi of scientific theories are essentially based on analogies with models drawn from familiar experiences. It seems valuable to attempt to uncover the analogies upon which a causal metaphysics is based. Some hints as to what such an analogy might involve are to be found in the suggestions made by some philosophers as to why they reject the explanatory adequacy of teleological patterns among events, for it appears that teleological patterns (patterns in which the logically antecedent conditions upon which the logical consequents of the pattern are to be inferred, occur temporally subsequently to those consequents) are rejected precisely because they can't be more familiarly understood in terms of the sorts of analogies that underlie the causal metaphysical model. For example, Scheffler simply assumes, with little or no attempt at justification, that we will wish to reject the notion of "final causes" as acting upon earlier events.[17] In regard to the explanation of an act by reference to its goal, he suggests that we are not to understand this reference as such as constituting an adequate explanation since the goal of a person, that for which he decides to do something, will--if it is achieved--occur later in time than the time of the decision, and hence

cannot causally explain the decision in question.[18] "Final causes", for Scheffler, cannot amount to explanations of decisions or actions because they cannot have (causally) acted upon these earlier events.

Braithwaite characterizes teleological explanation as a type of explanation in which there is specified a goal or end toward the attainment of which the explanandum event or activity is a means. He goes on to describe such explanations as puzzling because ". . . the explicandum is explained as being causally related either to a particular goal in the future or to a biological end which is as much future as present or past."[19] And this in spite of Braithwaite's own recognition that an explanatory law might with equal propriety assert a regular sequence, a regular simultaneity, or a regular precedence,[20] and that a law of regular succession can be reciprocal, i.e., such that the later event nomically determines the earlier event quite as much as the earlier determines the later.[21] (Thus the temporally asymmetrical causal relationship is accorded a scientific importance by even some neo-Humean philosophers who regard the relationship of nomological determination in accord with laws systematically interrelated as a whole theory as sufficient to account for the possibility of explaining and predicting matters of fact. The (extra) concept of causality involves, at least, the notion of one state of affairs necessarily determining another state of

affairs. Yet no thesis which is merely about the "necessity" with which we expect constancies of conjunction among phenomena can sustain the view that there is an element of necessity in the various lawful sequences of phenomena themselves. And this difficulty persists however systematically interrelated and logically necessary our expectations of constancy might be.[22])

In any event, given that a causal metaphysics is to be adopted, there are certain experiential models from everyday contexts which will clearly not do as interpretations for the deterministic relation. As pointed out by the University of California Associates of 1938 and also by A. K. Stout,[23] the causal relationship should not be construed as a relationship of "compulsion" of the effect by the cause. The notion of compulsion always implies, at least, the possible existence of desires in the consciousness of the entity said to be compelled. The notion of compulsion would appear to function as a useful explication of the relation of determination only if it could additionally be supposed that the entities which change are conscious organisms having desires.

As Schlick has pointed out, it is also not plausible to regard a causal relationship among events as if it were the result of the imposition of a law of nature upon reality, ". . . compelling nature to behave in a certain way, just as a civic law would force a certain

behaviour upon the citizens". Thus Schlick has deemed it necessary to explain that

> Laws of nature do not prescribe a certain order to the world, but simply describe the order of the world; they do not command what must happen, but simply formulate what does happen.[24]

Such an analogy would, again, seem to be inadequate in virtue of presupposing the existence of consciousness and intentional behaviour on the parts of the entities participating in the lawful sequences of events.

Schlick is also the source of another warning which has greater significance in the present context because of its concern with what appears to be a plausible analogy for the deterministic relation. After stating that determination means nothing but the possibility of calculation, Schlick cautions that "It does not mean that [the determining state] C in some magic way produces [the determined state] E".[25] Schlick explains that the calculation of E on the basis of relevant laws does "bring out" E, in the sense in which a mathematical computation, as a process of analysis, can only bring out that which is already contained in the presuppositions. Nevertheless, he argues, this ought not to lead to the use of the idea of "production" or "bringing forth" as an analogy for the internal relationship in the sequence of lawfully related events itself. William Kneale regards the sense of cause

to which Hume paid so much attention as being just this usage of the word 'cause' in which it is supposed to be an event which produces another event later in time but contiguous with the first.[26] And in the context of arguing that wants and beliefs may be causes, Alvin Goldman claims that ". . . the statement that S flipped the switch in order to turn on the light implies . . . that his having this want and his having this belief caused, or resulted in, his flipping the switch".[27] What is at work here is the notion of a cause as a source, or origin, or producer of its effect, or, perhaps, as the transmitter of an efficacy to produce an effect.

Richard Taylor finds that common speech, as used to express such homely causal relationships as that of bricks breaking windows, implicitly expresses the idea of "efficient causation" in the Aristotelian sense of this phrase which is that of an object or substance exerting a power and thereby necessitating a result by bringing it about or originating it.[28] In fact Taylor believes that ". . . if we did not already have precisely the idea of an efficient cause that once played such a significant role in philosophical thought, we would not even understand the concepts of causation with which modern philosophy has tried to supplant it . . .",[29] particularly the concept of invariant succession between states, processes, and events. This concept of efficient causation appears to be still

tacitly involved in causal metaphysical interpretations of the thesis of determinism.

Recall the earlier point that a sufficient specification of causal conditions and effected conditions should result in two sets of values, specified for all relevant variables of the appropriate system, each of which is both a necessary and a sufficient condition for inferring the other. Earlier, the view was discussed that what seems to be a temporal asymmetry in the relationship between the corresponding sets of conditions is really the result of an asymmetry in our practical apprehension of sequences of events, and a consequence of our limited abilities at controlling the world. Temporal asymmetry would, according to this view, be said to be thus introduced not into deterministic theories themselves, but into our application of them in our practical activities, so as to mirror the order in which event sequences can become temporally asymmetrical means-end sequences for us.

There is clearly something insightful in this pragmatic account but it is not sufficient, as formulated, to explicate all cases of sequences which we regard as temporally asymmetrical. There are sequences of events which we do not regard as having any action-relevance at all, yet which we always regard as having a fixed order. A volcano erupts spouting fiery lava in streams which run down the sides of the volcanic mountain gradually cooling

and solidifying along the way. It does not seem to be a mere matter of the action-relevance of this sequence that we have great difficulty in conceiving the real possibility (and not just picturing to ourselves the visual display) of cooled solidified lava rock distributed round a volcano as gradually heating, melting, rolling back up the sides of the volcanic mountain, and falling into the volcanic center which is at the approximate center of this miraculous activity. There is no obvious way in which we could use any of the events cited above as a means for producing the others, yet the events only fall into one natural temporal pattern, as far as we can see.

But more importantly, even with respect to sequences of events which have a practical relevance for us, the question may be raised as to why it can become a temporally asymmetrical means-end sequence for us only in a certain order. Despite the implications of purely epistemological determinism, it is clear that we are unable to regard sequences of events among which is one which constitutes a human action as being only incidentally asymmetrical sequences which could in principle occur in a temporally reverse direction.

Richard Taylor notes the fact that metaphysicians prior to Hume regarded the difference between cause and effect as consisting in the fact that ". . . the cause of anything was always something active, and its effect some

change in something that is passive.' Thus the sun has the power to warm a stone, but the stone has no power to make the sun shine; it is simply the passive recipient of a change wrought by the sun".[30] 'And the importance of time in this conception was that causes and activity could not work upon the past. Taylor claims, for example, that "For anything to be a cause it must act upon something and, as a matter of fact--indeed of metaphysical necessity--nothing past can be acted upon by anything".[31] In Taylor's view, our concept of cause is really still that of "efficient cause" in the classical Aristotelian sense, although perhaps diverging somewhat for we do not always require that a substance, or entity, have been the cause of something; events may do as well. The cause of the broken cup may be its fall from the table, which is an event. In such cases, since we cannot regard the causal event as being itself the origin of the power or activity which resulted in the cup's breaking, we are likely to assume tacitly that someone pushed or knocked the cup over or that some further event, such as an earthquake, preceded the cup's fall. The causal event of the cup's fall is thus regarded as a transmitter of some force, energy, power, or efficacy which is already "given" in the conditions prior to its own occurrence, and which carries forward a sequence of events, or changes, in which the energy may be said to be consumed or spent. And this process apparently cannot

be regarded by us as operating backward through time, conceived as the framework in which casual sequences are apprehended. Or perhaps our sense of "time's arrow" is the sense of the asymmetry of causal efficacy so that the concept of intrinsically asymmetrical causal efficacy itself constitutes the "time framework".

We have noticed that metaphysical hypotheses exceed our capabilities for experimental test. Yet it is no surprise that such hypotheses have frequently been devised. It seems quite puzzling that phenomena would exhibit such astonishing regularity if, as Hume thought, all things which are different distinguishable, and separable from each other in thought, and, perhaps, in fact, are things which may therefore exist separately and have no need of anything else to support their existence.[32] Such speculation may lead us to postulate on to those sequences whatever "glue"[33] is thought to be required in order that those sequences should hold together so uniformly and with such seeming non-contingency. And in the case of a good many regularities there are considerations which decidedly favour a causal metaphysical interpretation. If we tacitly presume ourselves to be agents possessing powers which enable us to act upon things, then the causal interpretation seems particularly capable of making sense of those sequences of events which transpire in experimental situations and in ordinary situations and which begin with the actions upon

things by human beings. The "quantity" of causal force or efficacy can be conceived as being "given" in the initial conditions, and as having originated in the experimenter or agent. We ordinarily take for granted the facts of our participation in many sequences of events which we regard as having been thereby altered in their course from what they would have been in the absence of our intervention. One possible model for the causal interpretation of lawful patterns of events thus seems to consist in the sorts of altered sequences which we regard as expending, through successive changes, a quantity of whatever "power" or "efficacy" we feel that we introduce into the world by our active intervention in the course of events going on around us.

But notice that there are really two significantly different sorts of sequences within the overall sequences which begin with the actions upon things by human beings. On the one hand we can think of sequences in which the initial event may be regarded as the initiation or origin of the causal activity, for example, the movement of my arm when with the cue I propel the billiard ball forward. On the other hand we can think of sequences in which the initial event may be regarded as merely the transmitter of causal activity, for example the movement of that billiard ball which leads to its impact with a second billiard ball, and so on. In the latter sort of case, the initial causal

event may be regarded as transmitting something of force, energy, power, or efficacy which the entities involved are unable to resist. The difference between these two sorts of sequence is important. The "cause" in the second case, namely, the movement of the first billiard ball, can itself quite readily be explained as the effect of some other cause; and this seems to result from the fact that it is not being regarded, in the context of its occurrence, as a true efficient cause, a true initiator of its own causal activity. This is not so clearly the case with regard to the first sort of sequence mentioned above. It is precisely such events as the human actions of shooting billiard balls which seem to involve the initiation of causal activity that are not so readily plausible or intelligible if they themselves are construed as "effects" on a causal model.

If these overall sequences provide any sort of model at all for a causal interpretation of regularities in the world, it would seem to consist in those sub-sequences involving collision and subsequent movement of the billiard balls, and which involve the mere transmission of efficacy or force. But such an analogy would be relevant to our understanding of events, if at all, only so long as we were not dealing with events which we ordinarily regard as involving initiation of causal activity. The collision analogy involves what may be called a "conservative sequence" and does not appear to be a plausible model for

understanding sequences in which there is an origination of something spontaneous or novel within the sequence.

Thus, serious problems would arise were we to attempt to extend such a model, involving the notion of a conservative sequence, to the interpretation of human actions in which, from a common sense point of view, there is the appearance of an origin or creation of power or efficacy. The problems are unavoidable so long as human behaviour is regarded as a proper object of scientific study. In ordinary contexts human actions seem to be thought of as creative in their own rights, and not merely as the passive recipients and transmitters of the causal efficacy which is already entirely given in conditions antecedent to their occurrence. "Transmitting" events may neutrally be regarded as either "causes" or "effects" depending on their temporal position within the narrower abstracted sequences that may concern us at any given time. But the creative events which are taken to initiate overall causal sequences cannot themselves be plausibly grasped as mere effects, even within sequences in which they are the temporally subsequent components.

Commonsensically, there is a world of difference between the initiation or creation of force and the transmission or expenditure of force. It may even be the case that the concepts of mere energy expenditure and of an energy-conserving system require the tacit assumption of

some prior act of energy creation. If this is so, then a metaphysics which involved only a model of conservative sequences, would not allow us to achieve a sufficiently complete understanding of the entire world of our experience. At the very least, given the prima-facie distinction between conservative and creative sequences, it is strange to think that either sort of sequence could serve as a metaphysical model by which to render the other sort of sequence more plausible or intelligible. We would not expect to gain greater understanding of a conservative sequence, for example, the rolling of a billard ball across a billiard table, by thinking of it as a sequence in which the ball at each succeeding moment propels itself anew. Expenditure of energy cannot itself be plausibly regarded as the perpetual creation of energy. Perhaps, then, we should not expect to be able to understand events which appear to involve spontaneity, creativity, novelty, initiative by thinking of them as the mere transmitters of force or efficacy which is already given in the conditions antecedent to their occurrence, and which can be regarded as "determining" them.

A great deal has thus far been said about the commonsense regard for familiar sequences. It has not been my intention to imply that there is anything particularly sacrosanct or inviolable about the commonsense world picture. But it does seem that it is this picture which

anchors our metaphysical interpretations of the purely deterministic structural patterns embodied in scientific theories, by underlying that which we shall regard as plausible and intelligible. We can at least hope that our commonsense is enlightened.

One potential source of such enlightenment will consist in resolution of the dilemma which appears to arise when the two distinct commonsense concepts of the nature of regular sequences--the conservative and the creative--are intermixed, one being used as a model for "understanding" the other. Several resolutions of this dilemma appear plausible. Firstly, we may wish to abandon metaphysics entirely as a serious intellectual pursuit. We would deal only with the purely structural (deterministic) patterns of our theories in the processes of explaining, predicting, and otherwise dealing "knowledgeably" with the world. Since my concern is with the <u>understanding</u>, not simply the explanation, of human behaviour, I have in effect chosen to speculate on the <u>nature</u> as well as the <u>structure</u> of deterministic behavioural sequences. I shall defend this choice only by referring again to two points noted earlier and derived ultimately from suggestions by Frank:[34] first, the untestable nature of current metaphysical views might perhaps be eliminated in the future through the operationalization of appropriate concepts; there is no reason to regard untestability at a given time as being

untestability in principle; second, metaphysical views may provide models, or "ideals", which human beings might profitably employ in directing the course of their lives.

As a second way of resolving the metaphysical dilemma, we may wish to demonstrate that relevant commonsense conceptions have not even been adequate commonsense. It might be thought that within the realm of pre-scientific commonsense, one sort of sequence can be shown to be, in effect, a variety of the other, that is, reducible to the other. Of greatest interest in this vein would be the attempts to show that the ordinary conception of freely willed acts was reducible to the ordinary notion of conservative causal sequences. Such a view might gain plausibility if it could be shown that ordinary moral concepts and notions of responsibility either did not require that distinct notions of free will be a part of the conception of human action, or were positively inconsistent with it. "Soft" determinism can be regarded as this sort of resolution of the metaphysical dilemma. A discussion of this view will be postponed until a subsequent chapter.

A third way of resolving the metaphysical dilemma would be to suggest an alternative model for a causal metaphysics. If the conservative model is not accepted, then the special apparently metaphysical meaning of the term 'causation' must be explained if it is really taken to mean something over and above simply "determination", as

explained previously in purely epistemological terms. Presumably, such a model would only be offered as a way of precluding the use of a "creative", or otherwise distinctively mentalistic, model for the understanding of human behaviour sequences. The intent would be to provide a causal model which was wide enough in scope so as to facilitate the understanding of both physical and so-called mental sequences of events. Thus the point of such a move would be to eliminate the need for a dual metaphysics for the purpose of scientific activity, regardless of whether or not a dual metaphysics was acknowledged to obtain at the level of commonsense.

A final alternative would be to simply accept the need for a dual metaphysics even at the level of scientific understanding. By choosing this alternative, we would be presuming a need for two distinct mutually non-reducible models for the metaphysical interpretation of the various deterministic regularities revealed by all our scientific theories. Hopefully this position would include a provision of some basis for distinguishing those regularities that were to be given one interpretation from those that were to be given the other interpretation.

A dualistic metaphysics would retain, for the understanding of human behavioural sequences and, perhaps, others, the conception of sequences involving creativity, spontaneity, initiative, novelty. It is occasionally

argued that this conception, which is essentially one of "efficient causation" in the Aristotelian sense, is based on our sense of ourselves as being such efficient causes. Thus, it is argued that we have some awareness of ourselves as active beings capable of just such initiative and creative acts as can lead to changes in the environment which need not, on the basis of given antecedent conditions, have otherwise occurred.[35] In order to avoid prejudging the question of the real duality of the commonsense conceptions of determination we might wish to modify this phenomenological, or introspective, starting point thus: we seem to experience our own personal initiation of action; we ordinarily think of ourselves as capable of spontaneously and creatively acting to bring about changes in the world. These ordinary notions serve as a kind of "data" which we may or may not wish to incorporate into our metaphysics. We might, for example, reject the metaphysical usefulness of such data if we thought that a metaphysics based upon them would impede the scientific activity for which that metaphysics was an adjunct, in particular, the scientific study of human behaviour.

However the primary point which seems clear in the light of this entire chapter is that an exhaustively causal, or mechanistic, metaphysics need not be true of the world in order that knowledge of that world be a real possibility for human beings. Thus an exhaustive causal metaphysics is

not a necessary condition for the possibility of human knowledge or the scientific enterprise; and the facts of the existence of human knowledge do not therefore exclude the legitimacy of a non-causal metaphysics. There is no justification for linking together, as, for example, Adolph Grünbaum has done, the concept of causality with those of predictability, lawful regularity, and the practice of the scientific method. Grünbaum has claimed that

> If human behaviour, both individual and social, does not exhibit cause-effect sequences, then the scientific method is essentially irrelevant to the elucidation of the nature of man, and both scientific psychology and the social sciences are permanently barred from achieving the status of sciences. This conclusion follows, since it is the essence of a scientific explanation in any field outside of pure mathematics to "explain" a past phenomenon or predict a future event by showing that these are instances of a certain law (or laws) and that their occurrence is attributable to the fact that the conditions for the applicability of the relevant law(s) were satisfied. Therefore, scientific or rational learning from past experience consists in ascertaining causal regularities from which to anticipate the future.[36]

If the word 'lawful' could be substituted for the terms which I have italicized in the quoted passage, then there would be no incompatibility between the view expressed by Grünbaum and the epistemological thesis of determinism which I endorsed in chapter two. And what follows from such a thesis is that a causal metaphysics is by no means necessary for the understanding to constitute a proper scientific study of human behaviour.

It is the intent behind this dissertation to show that what is ordinarily regarded as "freely willed" behaviour *is* knowable in terms of the lawful patterns which it exhibits. Consequently, if my arguments are correct, then freely willed behaviour, *understood as such*, possesses lawful regularities permitting prediction and explanation and scientific study. What alone shall differentiate this particular branch of science from others, especially physical sciences, will be nothing more--nor less--than the metaphysical interpretation to be put upon the nature of the deterministic sequences which human behaviour shall be found to exhibit.

If metaphysical hypotheses are, by definition, hypotheses which are formulated in terms that are, in principle, beyond the possibility of empirical test, then it may seriously be wondered whether they have any greater significance for us than the moral and personal comfort which I have suggested that they may provide. The hope that such hypotheses may some day be reformulated in operationally significant terms and hence become capable of test suggests that they are worth bothering with even at a time when they remain beyond possible test, but this point does not by itself suggest any means of discriminating those metaphysical hypotheses which are simply not *yet* testable from those which are *in principle* not testable. Thus, the legitimate question arises as to whether there is any

experiential difference at all that the acceptance of one among several conflicting metaphysical hypothese will make.

I have claimed that in order that we be able to accomplish the "goals" of science, namely explanation, prediction, and practical control, it is sufficient that the world be regularly patterned, or determined, and that our thought mirror in its regularity of inference pattern that deterministic structuring of the world itself. Thus, formal or structural relationships of world and of thought are the keys to the possibility of human knowledge of the world. But to say that a correspondence of the structuring of thought with that of the world is all that is required for the success of attempts to "know" the world is not to say that the "content" of the world cannot in some way be approximately thought. But if there is to be any sensible discrimination among metaphysical hypotheses about the non-formal aspects of the world, there must be some difference which is discernible within experience that will be made by one choice rather than another among competing hypotheses, and in particular there must be some criterion of successful or correct or right choice.

One possible criterion comes to mind. If metaphysical models were simple thought of as colourations superadded ex post facto to deterministic theories which were completed finalized accounts of various realms of study, then there could be no further experiential

difference that such models would make to the state of knowledge of, and the practical involvement with, the world. But as long as deterministic theories are in a state of incompleteness, then the very success at filling out our deterministic conceptions of the world might constitute one criterion of success or "rightness" for the methods or ways of thinking that inspired the insights by which the deterministic theories were rendered more complete. And a metaphysical model might surely constitute one way of thinking about a realm of phenomena which, if it were at all appropriate or fitting with regard to that realm of phenomena, could lead to a greater appreciation of the deterministic regularities involved in that realm. Thus metaphysical models might be "tested" by the overall heuristic fruitfulness which they bore as used to understand further the realms of phenomena for which they were being used as models.

III

The metaphysical view of human behaviour to be defended in this dissertation may be briefly outlined in the following way. Many of the significant behavioural sequences which we discern in human behaviour can be given the following very superficial schematic representation:

Wants + Beliefs (about present circumstances and about the actions which they make feasible for achieving those wants) ⟶ Action.

In accord with a causal interpretation, the action may be said to occur as a matter of causal necessity given the specified antecedent conditions. The antecedent conditions might be said to "bring about" or "result in" the action.[37] On such an interpretation the "wants" and "beliefs" of the above schema must be construed as mental states, dispositions, or events of the agent. It is these wantings and believings which, as such, "produce" or "give rise to" or "bring about" the performance of the action, either directly or by the mediation of an act of choice or volition which in turn "produces" or "gives rise to" or "brings about" the action.

Now ordinarily action is regarded as understandable once it is seen in the light of reasons for the performance of that action by that agent--"reasons" consisting in such things as what the agent wanted to achieve, what he believed circumstances to have been at the time of action, and so forth. Such reasons "rationalize" or justify actions by showing their appropriateness for the performing agent. And in ordinary contexts, such rationality or justification seems to operate as an explanatorily fundamental aspect of human action. Until we have appreciated an action in such terms, that is, until we see the "point" of an action, we

very often do not feel that we have an adequate understanding of why the action occurred.

I shall argue that the common sense concepts of free will and of rationality, in effect, dovetail. The non-causal model for the understanding of freely willed behaviour is what I shall call the "rationality model", a metaphysical interpretation in terms of the concept of rational (not causal) necessity. According to the rationality model, the "wants" and "beliefs" of the above schema are to be construed not as wantings and believings, that is, not as states, dispositions, or events--but rather as "that which is wanted" and "that which is believed". The wanted end "determines" the action in the logical sense of rendering it the means to achieving what is wanted in circumstances as these are believed to obtain. Thus what is wanted and what is believed together determine the rationality or appropriateness of a course of action. But that which is wanted and that which is believed do not causally necessitate the action which they logically rationalize. What is wanted and what is believed must be conjointly apprehended by a being capable of (practical) reasoning, and, furthermore, capable of initiating action for the sake of what he wants, on his appreciation of existing circumstances and on his appreciation of the actions which those circumstances make feasible for achieving what he wants. By initiating the appropriate

action, such a being may be said to act with "rational necessity", the necessity of willing the indispensably necessary means to one's given ends. It is this notion of "rational necessity" which provides us with the conceptual tools for understanding the freedom of the human will in a way which is non-causal yet compatible with the thesis of determinism.

A final question of importance should be noted in the context of this chapter. How might the viability of a dual metaphysics bear on the hypothesis of the unity of science? If the unity of science is defined in terms of the reduction of the terms and laws of the various sciences to the terms and laws of a single science[38], then the unity of science becomes a matter of the unity of mathematical calculi and the operational definitions and bridge laws linking these to empirical situations, across the various sciences. The possibility of such unity is not in any way threatened if a plurality of metaphysical interpretations are put upon that basic calculi in different scientific realms (though, of course, the need for such varying metaphysical models would require strong justification <u>given</u> the hypothesized identity of basic deterministic calculi).

On the other hand, if the unity of science is also thought to include an "intuitive" unification of the laws of various sciences[39], then the unity of science cannot in principle be achieved so long as a pluralistic metaphysical

outlook remains an essential aspect of the entire scientific attitude. For it is precisely the metaphysical unity of science that is challenged by the present dissertation. I have suggested that a metaphysical model for a particular scientific realm might be justified if it were heuristically fruitful in enriching the scientific knowledge of the world in that specific area. Thus a metaphysics of agency would be justified in the study of human behaviour if it opened up lines of thought to neglected or unappreciated deterministic patterns in the realm of human behaviour. If so, then so long as practical experiential success (in some inclusive sense) remains our final court of appeal, it would seem that the rewards of such a metaphysical outlook would more than compensate for the lack of "intuitive" unity to the entire scientific approach.

FOOTNOTES

[1]This point has been made before. Cf. Russell, op. cit., pp. 184, 118-189; Schlick, op. cit., p. 527.

[2]For examples of the neo-Humean viewpoint presented here, see R. B. Braithwaite, Scientific Explanation (New York, 1953), chs. 9, 10; Stephen Toulmin, The Philosophy of Science (London, 1953), ch. 5, esp. pp. 160-162; Nagel, op. cit., ch. 4; Carl Hempel, "Aspects of Scientific Explanation", § 2, in Aspects of Scientific Explanation (New York, 1965); N. R. Hanson, Patterns of Discovery (Cambridge, 1965), ch. 3.

[3]For discussions of such difficulties, cf. Russell, op. cit., esp. to p. 188; Schlick, op. cit.; Braithwaite, op. cit., ch. 9, esp. pp. 308-314; Toulmin, op. cit., pp. 119+; W. E. Johnson, Logic, Pt. III, chs. 4, 5; Richard Taylor, Action and Purpose (Englewood Cliffs, 1966), ch. 3; Herbert Feigl, "Notes on Causality", in Feigl and Brodbeck, ed., op. cit., pp. 408-418; Kneale, op. cit., Pt. II, § 15.

[4]This point was suggested to me by Professor C. A. Hooker.

[5]Examples of this formulation of the thesis are to be found in: Wilfrid Sellars, "Fatalism and Determinism", in Keith Lehrer, ed., Freedom and Determinism (New York, 1966), pp. 143-144; Goldman, op. cit., ch. 6, esp. pp. 172-173; Brand Blanshard, "The Case for Determinism", in Hook, ed., op. cit., pp. 3-4.

[6]Examples of this formulation of the thesis are to be found in: Adolph Grünbaum, "Causality and the Science of Human Behaviour", in Feigl and Brodbeck, eds., op. cit., esp. pp. 766-777; William Alston, "Wants, Actions, and Causal Explanation", in Hector-Neri Castaneda, ed., Intentionality, Minds, and Perception (Detroit, 1967), p. 302; Harold Ofstad, "Recent Work on the Free-Will

Problem", American Philosophical Quarterly, vol. 4 (1967), p. 181, and passim.; R. E. Hobart, "Free Will as Involving Determinism and Inconceivable Without It", in Bernard Berofsky, ed., Free Will and Determinism (New York, 1966), p. 79; A. K. Stout, "Free Will and Responsibility", in Wilfrid Sellars and John Hospers, eds., Readings in Ethical Theory (New York, 1952), esp. pp. 537-538; Israel Scheffler, Anatomy of Enquiry (New York, 1967), p. 89; University of California Associates, "The Freedom of the Will", in Feigl and Sellars, eds., op. cit.

[7]See discussion by Kneale, op. cit., p. 53.

[8]"Counterfactuals, Dispositions, and the Causal Modalities", in Minnesota Studies in the Philosophy of Science, vol. II (Minneapolis, 1958), p. 243. The entire discussion of Parts I and II, pp. 227-266, has influenced my subsequent remarks.

[9]Ibid., pp. 246-247.

[10]Discussions of this distinction are to be found throughout Frank, op. cit., cf. chs. 1, 2, 7, 9.

[11]For examples, see Frank, ibid.

[12]Ibid., pp. 10, 39.

[13]Cf. Alfred Jules Ayer, Language, Truth, and Logic (New York, 1935).

[14]Frank, op. cit., p. 37.

[15]Ibid., pp. 18-19. Cf. his discussion of the politically and morally motivated interpretations that have been put upon the theory of Relativity, pp. 172-188.

[16]Treatise on Human Nature, edited by L. A. Selby-Bigge (Oxford, 1888), p. 157.

[17]Op. cit., p. 113.

[18]Ibid., p. 89.

[19]Op. cit., p. 324.

[20]Ibid., pp. 308-309.

[21]Ibid., p. 312.

[22]The problem might be raised, with respect to this account, that by disavowing the existence of genuine (causal) necessity, the neo-Humean empiricist effectively denies himself the conceptual means for assuming that even those "systematic" regularities in our expectations of constancies in the conjunction of phenomena are themselves anything more than fortuitous regularities, i.e. than mere constant conjunctions in the associations among our ideas. Indeed within such a context, there is no sound basis for a distinction between genuine deductive inferences and just any regularly occurring transitions among judgments or beliefs. This problem has been recognized by Quine. Cf. W. V. Quine, Word and Object (Cambridge, 1960), p. 11.

[23]Cf. University of California Associates, op. cit., pp. 598, 600; A. K. Stout, op. cit., p. 537.

[24]Schlick, op. cit., p. 523.

[25]Ibid., p. 525.

[26]Op. cit., p. 53.

[27]Op. cit., p. 78. Italics are Goldman's.

[28]Op. cit., pp. 10-11, 19-21.

[29]Ibid., p. 19.

[30]Ibid., p. 32.

[31]Ibid., p. 35.

[32]Cf. op. cit., Bk. I, Pt. IV, § 5.

[33]A term used, derogatively, by Schlick, op. cit., p. 522.

[34]See footnotes #14 and #15.

[35]Cf. Taylor, op. cit., p. 13; G. F. Stout, Mind and Matter (Cambridge, 1931), p. 16; A. C. Ewing, Idealism (London, 1934), p. 172; C. A. Campbell, "Is Freewill a Pseudoproblem?", Mind, vol. LX (1951), pp. 463-464; R. M. Chisholm, "Freedom and Action", in Lehrer, ed., op. cit., p. 22.

[36]Grünbaum, "Causality and the Science of Human Behaviour", pp. 766-767. Italics my own.

[37]Cf. Goldman, op. cit., p. 78.

[38]Cf. Paul Oppenheim and Hilary Putnam, "Unity of Science as a Working Hypothesis", in Herbert Feigl, Michael Scriven, and Grover Maxwell, eds., Minnesota Studies in the Philosophy of Science, vol. II (Minneapolis: University of Minnesota Press, 1958), pp. 3-4.

[39]Ibid., p. 4.

IV

CAUSALITY AND MORAL RESPONSIBILITY

In chapter three (p. 112) I suggested that if the prima facie distinction between conservative (causal) sequences and systems and creative (free) sequences and systems could be shown to be inadequate even as "common-sense", then a metaphysics based on the use of this duality of models for understanding scientific regularities would be seriously undermined. It seems that so-called "soft determinism" can be regarded as an argument of this sort. "Soft determinism" can be understood as a thesis to the effect that the commonsense notion of free will does not really require a constituent notion of creativity, originality, or contra-causality, particularly in the realm of the most usual application of the commonsense notion of free will: namely, that of moral responsibility. In order to defend the sort of metaphysical duality which I am advancing in this dissertation, it is necessary, then, to show that soft determinism is not a successful thesis.

Furthermore, also in chapter three (p. 94), following Philipp Frank,[1] I suggested that one reason for engaging in metaphysical speculation concerning the nature of deterministic regularities, even though the hypothetical

results of such speculation were not capable of empirical confirmation or test, was that metaphysical hypotheses could become a model for human beings to use in living their lives, particularly their moral lives. Turning this idea around, we have the notion that a prior sense of the proper way for human beings to live out their lives, particularly in moral respects, can constitute one kind of consideration upon which to base the acceptance of a compatible metaphysics, or, at least, the rejection of an incompatible metaphysics.

This method of "justifying" a metaphysics may seem to be more rationalization, in the pejorative sense, than justification. But there does not seem to be any a priori restriction upon the type of argument or consideration which can serve in defense of a metaphysical hypothesis. Thus it seems in no way illegitimate that explicit moral commitments, commitments regarding the nature and presuppositions of moral judgments and moral sanctioning, might be made <u>prior</u> to any explicit commitments concerning the metaphysical nature of regularities in human behaviour, and might indeed constitute a basis for the latter. In effect, this consideration dovetails with that of the preceding paragraph. The moral commitments which provide the basis for assessing metaphysical accounts are precisely what must be analyzed in order to determine the extent to which commonsense involves dual notions of regular sequences.

The situation which I believe to obtain is the following: commonsense indeed involves a duality of notions of regular sequences; and this is particularly revealed in moral notions. The reconciliation of such moral notions with the thesis of causal determinism consequently cannot be effected. Of course this revelation, if the analysis shall be correct, does not amount to a thoroughgoing proof of the non-viability of causal determinism. This is because the very incompatibility of ordinary moral notions with causal determinism may persuade us to abandon those ordinary moral notions, thus exhibiting a prior commitment to causal determinism. The order of commitment appears to determine what our overall conclusions shall be. The most that can be shown is an incompatibility among various views to which we may legitimately commit ourselves in any order, restricted only by the canons of consistency. The important point to note is that since the truth of an exhaustively causal metaphysics seems not to be an a priori certainty and the arguments for it to be inconclusive, then it is no more reasonable to abandon moral commitments which prove to be incompatible with causal determinism than it is to abandon causal determinism on the ground of its incompatibility with those moral commitments.

The main topic of this chapter, then, concerns the viability of the thesis of (so-called) soft determinism.

I shall principally be concerned with the conditions for the justified ascription of moral responsibility. A possible source of confusion may be eliminated at the start. I shall not be analyzing the conditions which would justify the acts of ascribing moral responsibility or the behavioural acts of enforcing sanctions on ostensively moral grounds. Thus even if it can indeed be shown that the ascription of moral responsibility is incompatible with causal determinism, it does not follow that a commitment to causal determinism must be attended with any changes in that behaviour which is pertinent to ascribing moral responsibility. Thus it does not constitute an argument against causal determinism that causal determinists continue to ascribe moral responsibility and apply sanctions in respect of such ascriptions. For even if the thesis of causal determinism entails the falsity, or meaninglessness, of statements ascribing moral responsibility, assessing moral worth, and so on, it would not necessarily entail the unjustifiability of the linguistic acts of using such statements or the behavioural actions taken in support of them.

Furthermore, if the thesis of causal determinism indeed exhaustively accounted for human behaviour, then our acts of ascribing moral responsibility and of applying corresponding sanctions would be exhaustively caused to occur by the relevant antecedent conditions. Hence their

justification would be irrelevant to the facts of their occurrence. To suppose that a causal determinist can always alter his morally relevant behaviour simply on the basis of its inconsistency with mechanistic determinism is to ignore the very premise upon which causal determinism is based: namely that all events, including the behaviour of a causal determinist is exhaustively caused by relevant antecedent conditions. But the inconsistency among beliefs may not constitute a determining (causal) condition in every particular case. If causal determinism is true, then a causal determinist who holds his fellow human beings responsible for their actions and punishes them for their harmful actions, even with no intention to deter such actions in the future but only for the sake of sheer retribution, does so entirely because he is caused by antecedent conditions to act in this fashion. And this fact would constitute no argument against the truth of causal determinism.

Of course, the acceptance of the thesis of causal determinism may itself constitute a mental event which results in an altered overall mental state of such a nature that certain previous behavioural habits and beliefs will be terminated or altered as a consequence. But this need not invariably occur. And where it does occur, if at all, then there it would have been <u>caused</u> to occur (if causal determinism is true), and consequently the reasonableness

of such a change in beliefs can have played no essential role in this process.

With this point out of the way, I can turn my attention to the problem of the commonsense conditions for the justified ascription of moral responsibility. The most important consideration, for my purposes, is suggested by the familiar requirement that the agent in question, in order to be accounted morally responsible for his action, must have been able to do (or avoid doing) what he ought to have done (or avoided doing) even in circumstances in which he did not in fact do (or avoid doing) it. (I shall, from this point forward, regard an avoidance as a thing "done" and speak only about the agent's ability to "do".) Another way of saying that the agent was able to do what he ought to have done is to say that he could have done what he ought to have done. I shall use these phrases interchangeably. The term 'avoidable' can also be used to express the same condition: to say that an agent could have done what he ought to have done even in circumstances in which he did not in fact do it is to say that what he in fact did was avoidable for him in those circumstances. Finally, the condition can be expressed in terms of the "availability of alternative actions": if an agent could have avoided doing what he in fact did, then this is to say that there were at least two alternative actions available to the agent, each of which he could have done in the given circumstances.

The important thing to consider in the present context is the nature of this condition and its relationship to the thesis of causal determinism. Our assessment of the avoidability of an action for any given case is clearly going to depend upon a number of factors and circumstances which comprise that case. By speaking of the "availability of alternative actions", we focus attention, more explicitly than with any of the other formulations of this condition given above, upon the physical circumstances external to the agent and the ways in which these physically limit what can be done and also make certain actions physically possible by providing the wherewithal to accomplish those actions. Of course the external physical circumstances alone cannot be thought of as limitations or enablements without some further notions of the abilities, capacities, skills, and know-how of the agent. The agent is, at least in part, identified as a being of certain abilities, etc.; and these factors of the total situation are conjointly thought of as set over against the remaining circumstances comprising the external physical environment within which the abilities, etc. of the agent must operate. There is likely to be little or no disagreement from anyone, whatever may be their commitments regarding determinism, concerning the question of responsibility for action in those cases where, given the abilities, etc. of the agent, the impossibility of his doing other than what he actually

environment from which they nevertheless remain distinct. Non-human animals quickly come to mind; but plants will serve as an example as well. Changes in their sizes, shapes, and positions (that is, attitudes, rather than, in most cases, locations) are clearly often the result of their own natures as these respond to changes in surrounding environments. Even non-living purely physical systems fit the concept of a coherent, unified, active being in the above sense: for example, a sailing craft upon the water responds to the winds and the motions of the seas in accord with its geometrical properties and other features integral to its identity as a particular total "self". Of course, such systems do not appear to possess what we regard as a "mental life"; they do not appear to <u>desire</u> or <u>intend</u> the activities in which they become engaged. But a causal account of human behaviour might have to incorporate dispositional characterizations, in terms of tendencies to certain kinds of <u>overt</u> <u>activity</u>, of such notions as that of desire and motive. And if so, then there is going to be great difficulty in sufficiently explicating, along causal lines, the distinctions between activities which proceed from (hypothesized?) desires and motives, and those which proceed from other structural or internal features which are integral to the entities under consideration.

The distinction between action from want and intention, broadly speaking, and action from other dispositions

just beginning to drink; his drunkenness and consequent physical impairment could have been avoided had he drunk less. The situation may be extended further, however. The person may have an inability to control his general drinking habits; within this wider context his action may seem, once again, unavoidable. Nor is the decision about his responsibility necessarily over at this point. For we may, or may not, wish to regard his alcohol addiction itself as avoidable for him depending on the circumstances by which it came about, and, in particular, on the active role which the agent appears to have played in bringing the condition upon himself.

Also pertinent to deciding the responsibility for an action is the extent of foreknowledge, on the part of the agent, regarding the consequences of his action. We may feel that an action was physically avoidable for some person, yet feel that we cannot have expected the person to have avoided it on the ground that the detrimental consequences of the action which concern us were not reasonably foreseeable to the person from his point of view within the situation. We will think ignorant, rather than morally culpable, a person who gradually poisons his family by continuing to cook in a copper kettle the tin lining of which has worn away, but does so due to lack of knowledge regarding the harmful consequences of using such a utensil. Here, too, the context in which we place the action may be

extended. It may be that such ignorance of consequences could have been avoided by the agent, had he taken greater pains to be knowledgeable regarding his daily activities; so responsibility may be assigned after all, although at a different level.

These brief comments by no means exhaust the various considerations and qualifications which affect our commonsense assignment of responsibility for action. But they suffice to provide a setting in which we can begin to appreciate the problem which the thesis of causal determinism appears to pose with regard to the ordinary ascriptions of moral responsibility. Superficially the "moral problem" for causal determinism can be posed in the form of two questions: Why, given that human actions are regarded as causally necessitated no less than other physical events, nevertheless, are only <u>human beings</u> "held morally responsible" for the actions which they perform? Furthermore, since human beings are not held morally responsible for <u>all</u> the actions which they perform, how is a justifiable distinction drawn between actions for which they are, and actions for which they are not, to be held morally responsible?

If we took a hard and fast stand on causal determinism, then we might simply abandon our prior commonsense notions of moral responsibility. "Soft determinists" are those who have not wished to go quite so far, and have

typically responded to the "moral problem" by appealing to the following sorts of considerations. The concept of responsibility is designated by some to be a "practical concept" to which metaphysical notions of causality are viewed as irrelevant. This is to say that metaphysical notions of causality are compatible and consistent with ordinary concepts of responsibility by virtue of simply failing to deal with any of the same matters. Most typically,[2] the concept of responsibility is thus viewed as requiring freedom of the agent in the sense of an absence of (external) compulsion or constraint upon the individual which would have compelled him to act, or which would have constrained him from acting, in a way contrary to that which he wanted to do.[3] Compulsion and constraint are notions which are peculiarly applicable to conscious organisms, organisms characterized by desires and wants, and thus by tendencies to act in ways appropriate to achieving those desires and wants. Wants and desires may be regarded as features of the total "selves" of human agents. Compulsion and constraint are, then, basic sorts of obstacles to the fulfillment of the natural tendency of such agents to act in ways appropriate to themselves, that is, to their wants and desires. Thus, according to this view, agents prevented from acting in such ways by factors external to their "selves" are therefore not free in the morally relevant sense of this term.

The morally relevant sense of "freedom" is thus taken by soft determinists, in a negative sense, to signify an absence of compulsion and constraint. In a positive sense, it is variously expressed in the following ways: as the acting of one's <u>own</u> desires, desires which have their origin in the regularity of one's character and are not imposed by an external power;[4] as action which is "determined by" one's own desires, motives, and intentions rather than the desires and intentions of someone else;[5] as ability to act in response to the strongest motive at the time;[6] and as the power to will what one wants and desires, a process proceeding lawfully and without interference.[7] Notions of determination and of power and ability are problematic in the present context. It would seem that the most appropriate way to characterize the morally relevant sense of freedom espoused by soft determinists is that of the <u>conformity</u> of action to motive, want, or desire. Importantly also, this conformity should best be seen not as a mere conformity to single motives, wants, or desires of the performing agent, but rather, as in the view of A. K. Stout,[8] as a conformity to the self which is said to will the action. Again, the "self" is regarded as that complex whole within which a number of desires, motives, and so forth, interact, the operation of each ". . . determined more or less by the unity of the whole process within which they are factors; and the tendency of the self in the

whole process is towards its own satisfaction as a whole". R. E. Hobart claims that freedom is the property only of transient but unified beings, having a detached continuous mental life, ". . . a life reactive, but reacting according to its own nature".[9]

The important basis upon which notions of moral responsibility and moral freedom are to hinge are notions of a distinguishable "self" or morally relevant agent, one possessing unity and coherence as such. The morally relevant agent is an active being in continual interaction and involvement with a surrounding environment from which it nevertheless remains distinct. According to soft determinism, the identity of such a being encompasses, and is fully constituted by, such aspects of personality as desires, wants, motives, intentions. When these conditions are the proximate causes of the activities of the agent (with due allowance for acts of volition, if this be thought necessary as a part of the theory), assuming a favourable environment, then those activities are properly designated the "freely willed" actions of the agent. Moral responsibility for their consequences can then be incurred by the agent in respect of such activities.

However, at the level of common sense, there are thought to be more entities around than human beings alone which have coherence and unity as active beings in continual interaction and involvement with a surrounding

environment from which they nevertheless remain distinct. Non-human animals quickly come to mind; but plants will serve as an example as well. Changes in their sizes, shapes, and positions (that is, attitudes, rather than, in most cases, locations) are clearly often the result of their own natures as these respond to changes in surrounding environments. Even non-living purely physical systems fit the concept of a coherent, unified, active being in the above sense: for example, a sailing craft upon the water responds to the winds and the motions of the seas in accord with its geometrical properties and other features integral to its identity as a particular total "self". Of course, such systems do not appear to possess what we regard as a "mental life"; they do not appear to <u>desire</u> or <u>intend</u> the activities in which they become engaged. But a causal account of human behaviour might have to incorporate dispositional characterizations, in terms of tendencies to certain kinds of <u>overt</u> <u>activity</u>, of such notions as that of desire and motive. And if so, then there is going to be great difficulty in sufficiently explicating, along causal lines, the distinctions between activities which proceed from (hypothesized?) desires and motives, and those which proceed from other structural or internal features which are integral to the entities under consideration.

The distinction between action from want and intention, broadly speaking, and action from other dispositions

might be thought to be based on the following consideration. The former and not the latter shows a peculiar and distinctive susceptibility to change as the result of the practices of expressing moral judgments and applying appropriate sanctions in respect of them. Thus experience seems to show that such sanctions are most usefully and effectively applied to "behaving" entities in just those cases where the behaviour in question proceeded most proximately from such characteristics of that entity as desires, motives, and the like. Desires, motives, and the like might therefore be distinguishable from other traits, even on behavioural grounds, on the basis of the efficacy which we find ourselves to have in altering the future behaviour of the organism through the direct alteration of such characteristics (by means of argument, punishment, etc.).

Thus, Schlick has argued that the full significance of assignments of moral responsibility for actions is (commonsensically) taken to consist in the recognition that important factors present in the causal ancestry of these actions, such as motives and desires, are such as can be favourably affected in respect of future behaviour by either reward or punishment as appropriate.[10] Clearly the fact of the occurrence of an event of willing as an antecedent factor in the causation of an action is taken to indicate the causal efficacy, in the production of that action, of

something like a desire or motive on the part of the agent. And desires and motives are enduring features of human agents which are occasionally open to direct manipulation and control by other agents.

It is not, however, entirely clear that this consideration will enable us to distinguish among different kinds of dispositions in the requisite fashion. We can alter the future activities of other coherent unified systems as well, particularly by means of physical changes wrought upon them that might just as well be conceived of as a form of "punishment". A vending machine takes my quarter, yet refuses to hand over a can of soda pop; apparently it is empty, yet the coin return which should have worked in this case has failed. The vending machine has "stolen" my quarter, so, in anger, against this outrage, I "punish" the machine by kicking it. I find that very often, in these instances, such tactics will confront the machines with an immediate knowledge of the error of their ways and they will contritely give up the quarters which they had attempted to procure. (Sometimes, out of contrition, they will offer me a few extra quarters as added apology.)

Besides failing to satisfactorily ground the above distinction between dispositions which proceed from wants and desires, and those which don't, the consideration about amenability to change through our active intervention does

not in itself disclose any reason for regarding only the actions which proceed from these sorts of dispositions as capable of incurring moral responsibility for the agent. Furthermore, C. A. Campbell has argued that such an interpretation of the assignment of moral responsibility will simply not do for instances where we still assign moral responsibility for actions even though we are personally unable to affect the agent by our own actions, indeed, even where the motives and character of the agent may be completely beyond the possibility of any further alteration, as in the case of actions performed by agents now dead.[11] Furthermore, in the context of quite ordinary practical discourse, we are conceptually able to distinguish between the assignment of moral responsibility for an action and the recognition that, by sufficient reward or punishment, a specific sort of behaviour on the part of some individual can be reinforced or constrained. Indeed, we often attempt to reinforce or constrain behaviour even where there is no question of assigning moral responsibility for its performance, as, for example, in the case of young children whom we do not usually hold to be genuinely immoral even where the consequences of their actions are very harmful to others. Thus it is neither a sufficient, nor even a necessary condition of the assignment of moral responsibility to agents for their actions, that the assigner think himself effective in influencing the future behaviour of those

agents. Schlick's interpretation of what we mean when we assign moral responsibility for action is thus not an adequate account to cover all such assignments.

Nevertheless, it does not seem to be entirely wrong to think that there is something of moral relevance in the amenability of certain dispositions to future change by "morally relevant" methods--even, if we do not go so far as to think that all that we mean by assigning moral responsibility for action is that such action manifests this amenability. Argument, criticism, and praise are particularly important "morally relevant" methods for changing the future behaviour of agents; and it does seem that only such dispositions to action as result from desires, motives, and the like are amenable to change through the application of these techniques. I have never yet been able to induce a vending machine to return my "stolen" quarter to me merely by instructing it on the error of its ways. Thus there is something distinctive about dispositions to overt behaviour which result from desires, motives, etc., and part of the distinction appears to consist in amenability to change through techniques of linguistic manipulation, particularly those techniques which we should generally characterize as "appeals to reason" (rather than, say, insults).

Let us hold this point in abeyance for the time being. The "moral problem" for causal determinism is far from resolved, at this point. To see that this is so, we

the given situation). But careful reflection on the matter will show that the importance of circumstances, in our assignments of moral responsibility, extends beyond the mere physical limitations which they impose. Campbell refers to the common tendency to consider the degree of moral responsibility to be inversely affected by the degree of bad heredity and bad upbringing.[12] Techniques of brain-washing come to mind as an example of circumstances which exert psychological pressure upon individuals and in that way, may affect their attitudes and behaviour. Situations which generate extreme anguish for an agent may impel us to withhold attributions of moral responsibility for the actions performed by the agent thus affected. It thus seems that if the presence of an act of choice or will, or of corresponding wants and desires were the essential factor on the basis of which moral responsibility was regarded as being justifiably assigned, then heredity and upbringing would be totally irrelevant to the assignment of moral responsibility, for acts of choice and will as well as wants and desires occur no less frequently in the case of individuals with bad heredity and bad upbringing than they do in other cases. Thus certain conditions are regarded as imposing psychological limitations on the extent of alternative actions available to an agent in a situation characterized by those conditions. Such "psychological" factors may be of such an overpowering nature that

resistance to their pressure is the rare and exceptional thing, and therefore not something which could reasonably be demanded in most cases.

The situation which the agent faces must thus be regarded as including not only physical conditions external to the agent but, as well, certain conditions "internal" to the agent. And among these internal conditions will be not only the physical conditions of the agent's body but, as well, certain psychological factors which might also be of such a condition as to inhibit the agent's normal functioning. Thus the "external circumstances" on the basis of which responsibility may be excused for an agent at least partly comprise the very nature of the agent himself. The desire to live, to avoid great pain and to avoid harm to one's loved ones--these are ongoing characteristics of most human beings as agents. Such basic motivating factors are not mere capricious whims which it is open to human beings to contravene easily in their actions even when the disadvantages, on other grounds, become suitably high. Perhaps it is because such basic desires are not a matter of real choice, in any practical sense--and because we who judge also share in similar motivations--that moral responsibility for the actions which follow from them is often not ascribed even in cases where the actions also involve some other detriment which, in an abstract sense, it was also the agent's duty to avoid.

shall have to approach the issues from a new tack.

As suggested above, we commonsensically distinguish an agent from his external situation within the total interacting system, and then assess responsibility for the changes in that system by determining whether or not a significant source of the change arises with what we identify as the agent. This identification importantly encompasses the mental, or psychological, aspects of the agent--his wants, beliefs, fears, intentions, and so on. An individual is regarded as responsible for his activities if these activities can be shown, or assumed, to be traceable back to the mental aspects of the individual. The agent, as a certain definite being, is thus taken as given, albeit not fully known or understood. The various aspects of his personality are regarded not as independent factors in the process which leads to his decision to act in a certain way, but as aspects of a coherent unified whole being, a relatively independent sub-system within the entire system which comprises the agent interacting with an external environment.

It was noted earlier that the environment "external" to the agent was regarded as restricting the possible actions open to an agent, and that he will generally not be held responsible for actions which were physically impossible for him in the given situation (assuming that he could not reasonably be expected to have avoided being in

If these observations are correct, then it would follow that even within the agent himself as a complex unified whole "self", there is the possibility of something like compulsion and constraint, at least in the sense of something which constitutes grounds for waiving the attribution of moral responsibility for actions. But if we are willing to waive attributions of moral responsibility on the grounds of, say, the common human want of survival, then we must consider the possibility that for various individuals there may be certain uncommon wants which are of the order of strength that we attribute to the want of survival among people generally. Pervasiveness of a want seems to be no special defense in favour of according it moral importance; rather it merely explains such an accord as the result of our ready empathic appreciation of the agent's struggle to choose under such conditions. Consequently, we are led to entertain a consideration for specialized wants which would also compel (or impel) behaviour for the persons so wanting because they were not a matter of real choice in any practical sense.

True, there are limitations to the considerations that we will countenance under this rubric: it is difficult to see how we could ever excuse a fanatical philatelist from sacrificing his kingdom for a jolly good collection of stamps. A more plausible example might consist in the desire for revenge created in someone after he is made to

suffer unjustified agonies and humiliation. It might be thought that action proceeding from such motives should be exempted from ascriptions of moral responsibility on the grounds that such motives are not really consistent with the remaining wants and desires of the agents taken as unified wholes. They are not really "of" the selves which act under the force of their compulsion. This may be a correct account of how we identify compulsive wants, but it does not seem to be a sufficient account of why moral responsibility should not be ascribed in such cases. The reason for morally exempting the behaviour "compelled" by such motives seems to consist in the very compulsive or obsessive quality of the motives: as wants, they are unavoidable for the agent, and this remains an important consideration even though it is still true in those cases that the agent "could have acted otherwise" in the sense that he would have acted otherwise had he willed otherwise and he would have willed otherwise had this want not predominated.

Thus the difference between those cases of "wanted" behaviour for which moral responsibility is commonly ascribed and those for which it is not seems to be that in the latter sort of case, but not the former, the want is unavoidable for the agent. I suggest that the most plausible account of the commonsense significance of this factor is that it is regarded as diminishing the extent to which an agent is the creative originator of his own action

and augments the extent to which the overall motivation of the agent is causally necessitated ultimately by conditions external to the agent himself.

But on the grounds of causal determinism, the only distinction between the desires whose "obsessive-compulsive" nature mitigates the degree of moral responsibility for action and the ordinary desires which do not is the quantitative degree of forcefulness which characterizes the roles they play in the overall unified nature of the agent. And in accord with the causal model, even common ordinary non-compulsive wants and desires can themselves be analyzed as effects of conditions which are ultimately traceable to sources external to the agent. The notion of an agent in confrontation with an environment from which he yet remains distinct is beginning to dissolve before our very eyes. For it seems that none of the processes which the "agent" undergoes in the choosing or deciding how to act and in the willing of action can be regarded, qua causal processes, as essentially distinct or discontinuous from the processes of the surrounding environment.

It is precisely the notion of behavioural discontinuity with the surrounding environment that I should identify as "contra-causal" activity. If the above account is correct, then it would seem that, as, for example, Campbell has argued, the morally relevant commonsense notion of responsibility does indeed involve a notion of contra-

causal freedom.[13] But a central thesis of this dissertation is that causal determinism is not to be mistaken for determinism simpliciter. Thus, pace Campbell, to say that moral responsibility requires the possibility of contra-causal freedom is not to say that it requires the possibility of contra-deterministic freedom, or randomness. Rather it is to say that all antecedent conditions up to the time of the choice or will to act as the agent did, do not causally produce or causally necessitate the action which occurred. In terms of the notion of causal production described in the preceding chapter, this is to say that the will or choice to act was not a mere transmitter of a "quantity" of causal efficacy given over to it by the antecedent conditions, the wants and beliefs, and which had been given over to them by yet further antecedent conditions. For if this were the appropriate analysis, then there would be an unbroken causal sequence beginning not with the choice or will to behave in a certain way, but extending beyond this indefinitely backward in time. The choice or will to act would itself constitute no novel or original element in the causal sequence, so that there would be no particular reason to stop tracing the causes of action at this point and assign "responsibility" to the entity in which this event took place, than to trace the causes of action back to another link in the chain and assign the "responsibility" for the action to another entity at that point.

The most important "moral problem" for causal determinism is thus the way in which that theory undermines the notion of an independent "agent" upon which the ordinary attributions of moral responsibility have to hang. Since this is the key point against causal determinism as an account of human behaviour, it will be well to reiterate the basis upon which it is charged. As we have noted, the factors regarded as "external" to the agent, and therefore as excusable constraints and compulsions on his behaviour, may even be "internal" to the agent himself, internal even to his mind. Such factors might only be distinguishable as such within the very identity of the agent because they appear to involve patterns of activity which are at variance with the remaining aspects of the agent regarded as a unity, or perhaps because the agents themselves do not feel generally satisfied with their own activities in those respects and yet do not feel sufficiently "in control" of themselves as wholes, to prevent those activities from occurring. This sort of "feeling" might be familiar to us in such homely situations as that of avoiding what we <u>know</u> to be overeating and excessive drink.

Perhaps in such cases we can discern a history of events, finally rooted in some source external to the agent, which explains the incongruous behaviour by showing how it arose with circumstances not at all <u>of</u> the agent. Perhaps the agent is regarded as the passive recipient of some

externally originated pressure under the influence of which he could not have helped acquiring the habit or pattern of response which he now occasionally displays (under appropriate circumstances). I suggest that this notion of origin, of the power or efficacy which culminated in the present action, is the root of any feelings we have about attributing responsibility for that action. We seek some origin or source of what has happened. And there are some respects in which we are prepared to count the agent as "source" of this activity and some in which we regard the agent as constrained by factors not of his doing and which we suppose to be traceable back to points of origin ultimately external to the agent himself.

Whatever integrity or unity a human being has as a moral agent distinct from his surrounding environment, he has because, and to the extent that, his behaviour occurs in accord with an internal principle, or principles, of activity. I suggested earlier (p.139 f.) that moral agents must be "distinguishable selves", active beings having coherence and unity in the process of interacting with the surrounding environment. But I also noted that this characterization in itself is not sufficient to distinguish moral agents from integrated non-moral systems. A merely physical entity can remain distinguishable from its surrounding environment. This can occur because structural, or other internal or constituent (physical), properties, in

accord with physical laws, determine characteristic responses of that entity to given external conditions. But the laws which govern the internal properties of such merely physical entities are not metaphysically distinguished from the remainder of physical laws which govern the effect upon those entities of the environmental conditions. We may put this point by saying that such "internal" physical laws are not "emergent" with respect to physical theories; they are simply derivable from physical theories. There is no reason to interpret them metaphysically in accord with analogies that distinguish them from other physical laws.

If it were correct to speak of "selves" which were moral agents but were constituted in accord with the same physical principles that govern the entire universe, then there would be no reason not to anthropomorphize all entities of the natural order, human or non-human. But there seems to be a reason to interpret differently the laws which express the internal principle, or principles, governing the activities of human beings. This reason is the very moral relevance which such activity possesses in contrast to the moral irrelevance of the activities of purely physical entities. I have attempted to show that the moral concepts of commonsense precisely presuppose a differentiation of moral agents, namely, human beings, from non-moral entities and processes, particularly purely

physical entities and processes, and an irreducible discontinuity of human behaviour from the surrounding (morally irrelevant) physical environment. That is, the concepts of moral agency and moral responsibility presuppose the metaphysical distinctiveness of the moral agent. It is precisely the conceptualization of all human behaviour in terms of the same causal processes that pervade the physical environment surrounding human behaviour that would obliterate the discontinuity and render illegitimate the selective assignment of moral responsibility to only those human "selves". The irreducible discontinuity of human behaviour from the surrounding physical environment must be explicable in terms of the determination of human behaviour, at least in part, in accord with laws that are to be given a different metaphysical interpretation from the laws of physical science.

The view which I am recommending seems to be just the sort of view that David Armstrong characterizes as "Emergent-law" Materialism.[14] While it would not be a physico-chemical Materialism, still it would not, as such, seem to demand any emergent qualities, or an emergent substance such as "mind". Armstrong suggests that it would be natural at this point to hypothesize that these emergent laws were something that developed in all physical systems of a sufficiently complex degree of interrelation. In terms of the issues of the present chapter, this consideration

requires us to be prepared to ascribe moral responsibility to, say, sophisticated androids. I do not think that commonsense provides sufficient grounds for deciding how we should respond to such examples, although I, for one, would not reject the legitimacy of such ascriptions out of hand.

At any rate, a more important point with respect to Armstrong's view is that it seems to involve an equivocation on the term 'materialism'. This becomes clear when he suggests that a Materialist can <u>avoid</u> admitting the existence of emergent laws by simply arguing that although the whole range of human behaviour can be explained in terms of the working of the brain, nevertheless that working ". . . cannot be explained in terms of the physical principles that we <u>now</u> have".[15] But a Materialism which was not restricted to physical and chemical principles <u>as we know them</u> would not necessarily be a Materialism that Dualists are concerned to deny. Furthermore, if the principles of a "new physics", which encompassed the laws of the working of the brain, were <u>therefore</u> not derivable from the principles of the <u>old</u> physics, then there is a sense in which those new "physical" principles <u>would</u> be emergent with respect to the old physics.

But the important sort of "emergence" which I am concerned to link with moral conceptions is a metaphysical emergence. The principles which govern the "self"

determination of human behaviour must be interpreted in terms of different metaphysical categories than those used to understand physical processes. It does not matter whether the principles which govern this self-determination are those of current science or of future science. In either case, if they are metaphysically understood in terms of the same concepts used for understanding (non-moral) physical processes, then the resulting metaphysics is monistic and is precisely the sort of view which, I have argued, undermines the legitimate use of commonsense moral categories.

In accord with the dualistic metaphysics which I am endorsing, merely physical principles which would explain the workings of the external environment upon the human being, and would explain the physical dimensions, or aspects, of that being would be sufficient in themselves to explain, in all respects, ensuing human behaviour. They would be insufficient just to the extent that such behaviour genuinely represented or manifested determination in accord with the inherent nature of the human being. What is thus required for the use of moral language is a view of the self as, indeed, Hobart has characterized it, namely, a unified being having a detached continuous mental life, ". . . a life reactive, but reacting according to its own nature". The difference, in my account of the matter, is that such a concept must be construed contra-causally and in accord with

a process of self-determination, a process which I shall explicate more fully as one of rational self-determination, or, simply, rational determination.

A causal metaphysics, on the other hand, if applied consistently and thoroughly, seems to allow no room for any thoroughgoing distinction to be drawn between the morally essential nature of an agent and any other aspects of the agent, mental or physical--or even between the agent as a whole and his environment. What we commonsensically call a "person", a "human being"--the usual categories of moral agency--has no metaphysical importance in the face of causal determinism.* The application of such categories becomes quite arbitrary in terms of the consequences of that theory. "Persons" become the locations for sequences of events that we conceptually isolate from the surrounding environment only because they occur somewhat independently of it, enclosed within a "bag of skin".

Certainly there is no place within an exhaustively causal metaphysics for ascribing to those somewhat independent sequences of events, or the entity in which they occur, any genuine contribution as regards the overt activities of that entity. There would be only arbitrary beginnings to the patterns of human behaviour to which we pay attention--"antecedent" conditions beyond which we have simply ceased to search for further explanation. There would certainly be no originality involved in the activity

of such conditions. It is simply not adequate to characterize the causal process by saying, as Hobart has done, that "A man is a being with free will and responsibility; where this being came from . . . is another story. The past finished its functions in the business when it generated him as he is".[16]

If causal determinism were correct, then the regard for certain sequences of events as constituting whole integral morally responsible beings would be entirely without metaphysical foundation. A causal metaphysics appears to sustain no real concept of agency. But it is only agents, active creative beings, who can participate in moral responsibility, as this appears to be conceived by commonsense. Any apparent "agent" which we conceptualize would constitute, if causal determinism is true, a misconception based on inadequate knowledge regarding the true extent of the causal relationships between things; it would be the sheer result of ignorance. It is in this way that causal determinism entails either the erroneousness or the meaninglessness of the concept of an origin of activity which seems to underlie commonsense attributions of moral responsibility. From the commonsense point of view, it would make more sense to suppose that causal determinism could not constitute a complete metaphysical account of everything than to accept this account along with its moral implications. For one could only "reconcile" causal

determinism with one's moral values and judgments, that is, adopt "soft determinism", by imposing upon oneself veritable intellectual blinders and operating under a self-imposed naivety concerning the real nature and history of the events comprising human behaviour.

It seems that two conclusions of importance in the present context can be drawn from this consideration of commonsense moral notions. Firstly, it is quite appropriate, from the point of view of commonsense, to conceptualize the nature of various sequences of events in terms of (at least) two qualitatively different sorts of accounts: one, the passive, conservative, transmissive account; and the other, the active, creative, originative account. Only the latter sort of sequence has what might be called a real "beginning", the sort of event which alone cannot be regarded neutrally as either cause or effect. Thus commonsense really does provide the raw materials for a dual metaphysics.

Secondly, it is clear that a metaphysics of causality would entail the falsity, the unjustifiability and, depending on other views, perhaps the very meaninglessness of any judgments based upon the conceptions of genuinely creative original acts. If a commitment to moral values forms any part of our general conception of the right and good life for man, and if compatibility with this conception is to constitute one basis for the assessment of

metaphysical accounts, then a commitment to moral values entails the inadequacy of an exhaustively causal metaphysics, particularly the inadequacy of a causal account purporting to account for all aspects of human behaviour.

The sort of metaphysics which most readily promises to make sense of commonsense moral conceptions is a metaphysics of agency, of self-determining activity. In another chapter I shall be explicating, as fully as possible within the limits of this dissertation, this alternative metaphysical account of (at least some of) the deterministic regularities in human behavioural sequences. To round off the present chapter, certain points in regard to the moral issues remain to be clarified. Firstly, it may be recalled that earlier I suggested that an important part of what is distinctive about dispositions to overt behaviour which result from desires, motives, etc., appears to consist in the amenability of the organisms so characterized to future behavioural changes in those respects through the application of techniques of linguistic manipulation, particularly those techniques which we should generally characterize as "appeals to reason". Thus, what is of particular moral relevance in regard to the nature of an agent is, briefly, the rationality of that agent. This point has been amply developed by, for example, Kant and has recently been discussed by such philosophers as Kurt Baier, A. C. McIntyre and Sidney Hook.[17]

Hook notes that we attribute moral responsibility where there is a tendency to respond to valid reasons, to behave rationally, to respond to human emotions in a human way.[18] Baier characterizes the significance of rationality for the concept of moral responsibility by noting that bringing people to account and ascribing responsibility to them is an effective and justifiable practice only for beings whose behaviour can be regulated by the promulgation of social rules, people who can <u>follow</u> rules, orders, directives--that is, people who can understand the action-descriptions contained in such directives and can, at will, behave in ways to which such action-descriptions would apply.[19] Part of the significance of saying that an agent has an ability, on a particular occasion, to choose differently than he in fact did, consists in highlighting the general ability of the agent to recognize and weigh alternatives.[20] MacIntyre argues that the area of human freedom is to be made sense of by means of the concept of "rational behaviour", construing rational behaviour as behaviour which can be influenced or inhibited by the adducing of some logically relevant consideration.[21] In that the behaviour is thus "determined" by logically relevant considerations, there are necessary conditions for its occurrence. Such behaviour is, furthermore, often predictable because it often occurs rather consistently on the basis of principles, and such principles can therefore form

the basis for the prediction of such behaviour. (E.g., "She will never accept that bribe. She is a woman of great integrity".) But this predictability, MacIntyre warns, has nothing to do with causal determination. The agent's decision, conclusion, or deed is uncaused in that only the adducing of logically relevant considerations played a part in determining them.

One point on which, for example, MacIntyre is not clear, and which I hope to clarify in my own account, is whether or not he construes the preferences and motivations of the agent as causally efficacious conditions. He suggests that a complete causal account can be given of rational behaviour, although he qualifies this by saying that such an account seems to misdescribe rational behaviour, among other things, by blurring the contrast between rational appraisal and irrational reaction.[22] This vacillation in the use of the term 'cause' seems to occur because no fully articulated metaphysical explication of it is developed. It appears at times to mean simple determination in the epistemological sense in which I have construed this term, namely, that of calculability in accord with a lawful pattern. In elaborating my own account of the concept of rationality and rational necessity as a metaphysical model for freedom of the will, I hope to show that it is compatible with (simple) determinism in this sense, even though it is inconsistent with any metaphysical

notion of causality and requires instead the concept of an active creative agent. As I see it, the metaphysics of rationality becomes a metaphysics of agency.

Finally, a difficulty which pertains to the metaphysics of agency, in regard to the basis for assigning moral responsibility, must be noted. Earlier, I suggested (p. 146 ff.) that the presence of an act of choice or will, or of the corresponding want or desire, did not seem to be the essential factor on the basis of which moral responsibility was regarded as being justifiably assigned because such factors might be regarded as being entirely the result of external conditions and hence, to that extent, not the original autonomous activity of the agent himself. This analysis was taken to suggest that origination or autonomy is the key factor in the assignment of moral responsibility. On the basis of this analysis, I have introduced, and shall subsequently explicate, a metaphysics for interpreting human behaviour which seems to preserve this concept in the understanding of that behaviour.

The difficulty is that if an originative, or rational, interpretation is to be given indiscriminately to all behaviour sequences, then I have not established within my metaphysical framework a distinction which uniformly parallels the distinction between instances of commonsense ascription of moral responsibility and instances of commonsense withholding of such ascription. For, commonsensically,

we make moral distinctions among behavioural sequences, precisely on the basis of whether or not we are able to trace the origin of the want factor to an external causal source. Some (though only some) human actions are thus conceptualized as the causal consequences of external conditions, in that the wants from which they stemmed are themselves seen as merely mediate links in causal chains that are traced to sources external to the behaving organisms. The framework which I have proposed, to account for the moral autonomy of human beings preserves the "caused versus uncaused" dichotomy of commonsense, but at the expense of encompassing among autonomous original behaviour many instances in which, commonsensically, attributions of moral responsibility would be withheld.

The following modification of my view is possible: the rational framework could be regarded as applying to human behaviour in general just so long as no positive evidence in particular cases suggested that the wants in question were the causal consequences of external conditions. This suggestion seems unfortunately ad hoc, particularly since it is not based upon a fundamental metaphysical distinction that has grounds in the metaphysics of agency itself, but is rather based upon an epistemological consideration that may simply be an historical and incidental feature of a given state of knowledge (or ignorance) regarding human behaviour. However there are no solid

metaphysical considerations which I can suggest at this point to ground this suggested modification.

It may still be said, in defense of the metaphysics of agency as an overall framework for understanding human behaviour, that at least some basis is provided for distinguishing selves, that is moral agents, from their environments, that is, from physical entities and processes. The rational framework provides for the discrimination of coherent selves according to the presumption of inherent rationality (to be explicated in chapter nine) which defines the distinctive form of regularities to be found in the behaviour of such organisms to the extent that this behaviour is determined by the inherent nature of the organisms themselves. And from the moral point of view, this remains one advantage which a rational metaphysics has over a causal metaphysics.

FOOTNOTES

[1]Philosophy of Science (Englewood Cliffs, 1957), pp. 18-19.

[2]Some untypical examples are: Harry G. Frankfurt, "Freedom of the Will and the Concept of a Person", The Journal of Philosophy, vol. LXVIII, pp. 5-20; Robert J. Richman, "Responsibility and the Causation of Action", American Philosophical Quarterly, vol. 6, pp. 186-197; and Wilfrid Sellars, "Fatalism and Determinism", in Keith Lehrer, ed., Freedom and Determinism (New York, 1966), pp. 141-174.

[3]Cf. Moritz Schlick, Problems of Ethics (New York, 1939), pp. 147-153; R. E. Hobart, "Free Will as Involving Determinism and Inconceivable Without It", in Bernard Berofsky, ed., Free Will and Determinism (New York, 1966), pp. 78-79; A. K. Stout, "Free Will and Responsibility", in Wilfrid Sellars and John Hospers, eds., Readings in Ethical Theory (New York, 1952), pp. 537-538; University of California Associates, "The Freedom of the Will", in Herbert Feigl and Wilfrid Sellars, eds., Readings in Philosophical Analysis (New York, 1949), pp. 598-604, 613; Kurt Baier, "Responsibility and Action", in Myles Brand, ed., The Nature of Human Action (Glenview, Ill., 1970), p. 112; and Adolph Grünbaum, "Free Will and the Laws of Human Behavior", American Philosophical Quarterly, vol. 8, pp. 302-309.

[4]Schlick, op. cit., p. 154.

[5]University of California Associates, op. cit., pp. 599-600.

[6]Grünbaum, op. cit., p. 305.

[7]Hobart, op. cit., pp. 72-73.

beliefs and wants an essential explanatory relevance. What is relevant to the understanding of human behaviour in terms of its rationality, are thus primarily propositional contents of beliefs and wants, rather than believings and wantings. Thus the relevant rationally explanatory factors are primarily the universal intentional properties in virtue of which mental events, states, or dispositions constitute the having of precisely those beliefs and wants, rather than some others instead; they are not the mere havings of those mental events, states, or dispositions.

It is true however that the appropriateness of some action is not explanatorily relevant to the occurrence of the action unless the appropriateness of the action in some way participated in bringing about its occurrence. Yet there is no a priori reason to think that the only manner in which one thing can participate in bringing anything else about is through causal efficacy. This narrowed conception of how alone things can come to be is unfortunately further reinforced by such considerations as that raised by Gustav Bergmann regarding states of mind.[6] A state of mind, claims Bergmann, is a fact that "intends" another fact. A present state of mind which is a purposing (one sort of "reason" for action) intends a future fact which, because it may or may not come to pass regardless of whether the action appropriate to its obtainment in some situation occurs or not, can surely have no relationship of

[19]Op. cit., p. 113. It should be noted that Baier's account is not a non-causal account; he appears to believe that the "agent-causation" involved in rational activity can be accounted for adequately as a variety of ordinary event-causation in which the important causal condition of the agent's activity is a certain goal-directed operative propensity on the agent's part; pp. 110-111.

[20]Ibid., p. 112.

[21]Op. cit., p. 248.

[22]Ibid., pp. 252-253.

V

CAUSALITY, RATIONALITY, AND AGENCY

In the previous chapter, it was suggested that human behavioural regularities can neutrally be regarded as comprising the following:

Wants + Beliefs —— Action

The present chapter shall center on a discussion of the causal versus the rational interpretation of such sequences. In the first part of the chapter, I shall attempt to show that a rational interpretation of the above behavioural regularities is indeed in contradiction with a metaphysical interpretation of those regularities in terms of causality. In the second part of this chapter, I shall argue that there are important and compelling grounds for favouring a rational interpretation over a causal interpretation as a metaphysical account of human behaviour, in particular, a rational interpretation which involves a metaphysics of agency.

I

In accord with a causal interpretation of the above sequences, the action may be said to occur with

causal necessity given the occurrence of the specified antecedent conditions, namely, the wants and beliefs on the part of the agent. Causal interpretations for human behaviour have been variously formulated[1] but the central argument common to most of the formulations begins with the following sort of argument. It is acknowledged that there are a variety of common everyday explanations in which human behaviour is "explained" to the satisfaction of common-sense by appeal to the "reasons", "purposes", "beliefs", "motives", "intentions", and the like, of the performing agent. But from a more enlightened or a scientific point of view it is not thought to be clear just what sort of explanatory factors such mentalistic phenomena are.[2] Ordinarily these phenomena--events, states, and dispositions of mind--are regarded as explaining such behaviour by providing a rationale or justification for the behaviour: by revealing "the point" of the behaviour in question. However rationalization and justification are not thought to constitute explanatorily relevant factors at a more enlightened level of understanding. Thus Donald Davidson has suggested that ". . . a person can have a reason for an action, and perform the action, and yet this reason not be the reason why he did it. Central to the relation between a reason and an action it explains is the idea that the agent performed the action because he had the reason".[3] Similarly Alvin Goldman has pointed out that it is not "

adequate to explain an action of an agent simply by citing some end that the agent wanted to achieve and the agent's belief that by so acting, he would achieve his end, because an agent can have such a want and such a belief, yet have performed the action for some other reason, or not for any reason at all, i.e., accidentally. In order that such an explanation be adequate the reference to the want and the belief must also imply ". . . that his having this want and his having this belief caused, or resulted in, . . ." the action which we seek to explain.[4]

It is true of course that we cannot explain an action in terms of its rationality by appealing to factors which simply make that action rational yet have no further bearing upon its occurrence. But it does not follow from this that the relationship between the reason which really explains the action--the real reason why the agent behaved as he did--and the action which was performed must therefore be a causal relationship. I shall be arguing, in chapter seven, that when it is judged that a certain type of action is rational in the context of certain specified sorts of circumstances, given a specified type of wanted end-state, the rationality being thereby attributed to the action is purely a matter of the logical relationship, the relationship of appropriateness, which the action bears to the specified conditions. This may be tied in to certain ideas suggested by Karl Popper in Objective Knowledge[5], by

regarding an action to be rational if it corresponds to the solution to a "problem-situation" confronting the agent, namely the practical problem of achieving wanted ends. The "logical relationship" which the solution to a practical problem bears to that which is wanted and that which is believed can be represented by means of a practical syllogism according to which that action is implied, as the appropriate thing to be done, by the premises which state what is wanted and what circumstances obtain which are relevant to, and enable, the attainment of this want.

Most importantly, then, the appropriateness of action is a matter of the relationship between a goal in the sense of a possible state of affairs the realization of which can be wanted by someone, and a set of circumstances, an envisaged state of affairs, the actuality of which can be believed by someone. Either of these factors can be described by means of a statement expressing an appropriate proposition, namely the proposition which best characterizes the content or object of the belief or want in question. Believings and wantings may be characterized as attitudes toward, or ways of apprehending and regarding, those contents or objects, as expressed by means of the appropriate statements. The model for the understanding of human behaviour in terms of its rationality, as claimed previously, must be one which accords the logical relationship between the propositional contents, or objects, of

beliefs and wants an essential explanatory relevance. What is relevant to the understanding of human behaviour in terms of its rationality, are thus primarily propositional contents of beliefs and wants, rather than believings and wantings. Thus the relevant rationally explanatory factors are primarily the universal intentional properties in virtue of which mental events, states, or dispositions constitute the having of precisely those beliefs and wants, rather than some others instead; they are not the mere havings of those mental events, states, or dispositions.

It is true however that the appropriateness of some action is not explanatorily relevant to the occurrence of the action unless the appropriateness of the action in some way participated in bringing about its occurrence. Yet there is no a priori reason to think that the only manner in which one thing can participate in bringing anything else about is through causal efficacy. This narrowed conception of how alone things can come to be is unfortunately further reinforced by such considerations as that raised by Gustav Bergmann regarding states of mind.[6] A state of mind, claims Bergmann, is a fact that "intends" another fact. A present state of mind which is a purposing (one sort of "reason" for action) intends a future fact which, because it may or may not come to pass regardless of whether the action appropriate to its obtainment in some situation occurs or not, can surely have no relationship of

antecedent causal determination to that action. However the present mental state of purposing is completely and unproblematically a present fact. And so, for Bergmann, the having of a purpose qua present state of the organism can stand in relationships of causal determination to subsequent actions.

This is not to say that the propositional content which differentiates one particular event, state, or disposition which is a reason, purpose, etc. from others, would be of no significance whatever on a causal interpretation of behavioural regularities. This is far from necessarily being the case, even if a commitment to causal determinism is regarded as devolving into a thorough-going commitment to materialism. For even a materialist must acknowledge that differences exist among various so-called "mentalistic" phenomena such that different kinds of these can be recognized to be causally connected with different kinds of overt behaviour. In accord with materialism, the differences among such "mentalistic" phenomena would, of course, ultimately consist in differences in material properties and relations. So the so-called "mentalistic" differences, the propositional contents of various reasons for action, must be seen from a materialistic point of view as having material counterparts or embodiments, or as being redescribable, reconceptualizable, materialistically. For if this were not the case, then they would have to be

regarded as totally lacking causal efficacy, in short, as being epiphenomenal. But if so-called "mentalistic" phenomena could be reconceptualized materialistically, then nothing would prevent the propositional contents, qua material properties, from constituting causally efficacious aspects of those material phenomena which they would characterize.

Even for non-exclusively materialistic causal views, recognizable differences still have to be acknowledged to exist among various mental phenomena in order that different such phenomena can be apprehended in terms of causal connections with different kinds of overt bodily actions. From the causal point of view it may be the case that the only causally relevant psychological factors are, in Paul Meehl's terms, the "holdings" or "tokenings" of reasons, that is, the believings or wantings, because, as Meehl notes, these psychological factors alone are "in the world" and hence part of the chain of causality.[7] Nevertheless it must still be the case that the full causally relevant character of such states requires reference to the propositional contents of such "holdings" or "tokenings" in order that the logical relationships of rationality and justification--which intuitively seem to have something to do with explaining behaviour--may be translated into causal psychological relations. According to Meehl, it is possible to formulate causal laws of behaviour according to which

the "tokening" of reasons is shown to (causally) influence behaviour. And a critical element in that which makes certain kinds of mental events causally efficacious is that they are tokenings of sentences which, in the realm of logic, constitute what Meehl calls "valid reasons".[8]

A minor point about "reasons" should be clarified. The concept of a "reason" is largely relational in character, and is conceived in terms of its function within wider mental sequences of "reasoning". This point holds for both believings and wantings and the propositional contents of those believings and wantings. No mental phenomenon (event, state, or disposition) nor any feature of such a phenomenon is in and of itself a "reason", in either sense. A reason is that which fulfills a certain sort of role in a sequence of mental phenomena that constitutes a process of reasoning. A mental phenomenon constitutes a reason, in the sense of a believing or a wanting by virtue of its playing the role of the assertion of a premise in an actual process of reasoning, whether logically valid or not, which is gone through in someone's mind. A propositional content of a mental event, state, or disposition is only properly considered a reason when considered in its logical relationships to propositional contents which can differentiate further mental phenomena, whether or not they are gone through in someone's mind.

On the assumption that reasoning of a "means-end"

variety is characteristic of, and essential to, any pattern of thought which culminates in overt rational action, we are able to identify (at least) three basic sorts of mental phenomena which can function as the having, holding, or tokening, of reasons in practical, or action-relevant, patterns of thought. These are: (1) goals, purposes, principles, interests, wants, desires, or anything which, in a very wide sense, is an aim or end of action; (2) beliefs about the condition of circumstances which obtain at the time of the performance of action, and which may also pertain to further particular circumstances either prior or posterior to this time as is additionally thought relevant to the decision as to how to act at the given time; and (3) general cause-effect beliefs according to which ends of particular types and circumstances of particular types are regarded as having a special sort of connection with actions of particular types, a "connection" which consists basically in the general attainment of the type of end specified in the type of circumstance specified given the performance of the specified type of action. Thus we may say that a believing or wanting on the part of an agent, becomes a reason for that agent's acting in a particular fashion at some particular time, by being a mental phenomenon the propositional content of which is taken by that agent to constitute a reason in the logical sense for performing the action at the given time.

Thus an actual logically sufficient process of practical reasoning on the part of a person must involve (or assume) specification of each of the following: goals, details of the action-relevant situation, and general cause-effect relationships relevant thereto. When in ordinary discussion, we are asked "the reason" for our action, the context of such questioning may make it appropriate to selectively cite only one such factor as constituting the reason for acting. However no one such factor in and by itself ever constitutes a reason for acting.[9] This is the relational aspect of the concept of reason, its functional role within some wider sequence which involves a set of premises sufficient as a whole to imply or justify a certain conclusion. Neither a want nor a belief alone imply the appropriateness of action; but each can be a "reason" for action, given a sufficient specification of the other component. The argument that "reasons are causes" must thus amount to the thesis that behaviour is caused by a set of mental conditions--wants + beliefs--which causally determine behaviour. Consequently, when the term 'reason' is used in the present context it shall be understood either in the sense of an entire mental state comprising relevant wantings and believings, as in Davidson's usage of the terms 'primary' and 'secondary' reasons[10] or, in the logical sense of such propositional contents of wantings and believings as comprise a practical syllogism. The

context will make clear which sense is intended at any particular time.

As previously noted, at the level of commonsense, behaviour is ordinarily regarded as understandable once it is seen in the light of "reasons" for the performance of that action by that agent--though the sense in which "reason" is to be understood is not clearly defined as either propositional content or as believing or wanting at the level of commonsense. Such "reasons", albeit in this ambiguous sense, rationalize or justify actions by showing their appropriateness for the performing agent. In ordinary contexts such rationality or justification seems to operate as explanatorily fundamental aspects of the human actions which they explain. Until we appreciate action by seeing its "point", we very often do not feel that we have an adequate comprehension of why the action occurred. In few other contexts beyond that of human behaviour do we require such a condition to be met by explanations.

Thus in the ordinary context of explaining human behaviour, we, in effect, often regard the rationality of the explanandum as constituting an essential factor somehow determining its occurrence. Rationality constitutes a basic non-eliminable explanatory factor at the level of the commonsense understanding of human behaviour. And the rationality of action, as argued earlier, consists in the logical relationship which a description of that action

3 5

OF/DE

bears to the propositional contents of a given set of wantings and believings--a relationship which can be expressed by a corresponding practical inference in which that which is wanted and that which is believed are shown to entail a resolve to perform that action. However, as noted previously, logical relationships do not bear causal efficacy as such.

It might seem plausible to say that it is the wantings and believings themselves on the part of an agent that determine which action(s) will be rational for that agent. However though such a statement is not inaccurate as it stands, it is far from constituting all that needs to be said about this point. The wanting and believing on the part of a given agent determine which action(s) will be rational for that agent in particular circumstances precisely and only because the propositional contents <u>of</u> those wantings and believings bear a particular logical relationship to the relevant description of action(s) in question, namely, the relationship of determining the appropriateness of that action (or actions) as thus described. The causal model in itself, unsupplemented by any additional assumptions, provides no reason to think that wantings and believings should anymore give rise to actions which are <u>rational</u> in terms of what is wanted or believed than to actions which are <u>irrational</u> on those grounds. If action is construed on the causal model, then, without any

further presuppositions about the nature of human action, any relationship of rationality which it happens to bear to the wantings and believings which constitute its antecedent conditions must be regarded as a matter of pure happy coincidence. As is the case in respect to purely physical explanations, such rationality could not be regarded as having any particular determinative role or any explanatory relevance.

Surely this result conflicts with the way in which we commonsensically operate in our understanding of human behaviour. But perhaps the causal model as an interpretation of human behaviour can be supplemented so as to take care of this difficulty. The relationship of (practical) rationality, it can be agreed, is constituted by the logical relationship between the propositional contents of beliefs and wants and the propositional contents of appropriate action descriptions, descriptions which express what may be called, following Davidson, the "quasi-intentional" aspect of human actions.[11] However this formal relationship, which constitutes the rationality of those actions does not preclude the possibility that believings and wantings which, in virtue of their intentional features, happen to render actions appropriate can also stand in relations of causality to the ensuing actions. Mental phenomena which, in respect of their propositional contents, stand in such logical relationships to ensuing action, under the relevant

description, are not thereby precluded from being causally efficacious (within the contexts of appropriate background conditions) in the occurrence of those ensuing actions. Stephen Toulmin, for example, puts the point by saying that a statement of our reasons for behaving as we do is never equivalent to or inconsistent with a statement of the causes determining that behaviour, that they simply bypass one another by operating on quite different levels.[12] And Charles Landesman claims[13] that a teleological explanation of the form

> S did y because he wanted x and believed his doing y would get him x

is logically compatible with such causal statements as

> When S began to want x and to believe that his doing y would get him x, his brain was in a certain state T, and T was a causal condition of a further brain state U, and U caused his limbs to move in a certain way, and his limbs moving in this way was a precondition of his doing y.

Goldman goes a step further by asserting that it is a logical truth about wants that they tend to cause acts, that the concept of want includes the idea of tending to cause acts, although the specific acts which a given want has a tendency to produce depend on factors other than the want itself, specifically on the agent's beliefs.[14] Goldman does qualify this by saying that the logical truth that wants

tend to cause acts does not entail it to be a logical truth that some particular act-token was caused by some particular want; the latter is an empirical matter.[15]

Nevertheless this overall viewpoint is strikingly inconsistent with recent attempts to argue that a "logical relationship" between mental phenomena, including volitions, and overt actions precludes the possibility that the former can stand in relations of causality to the latter. From the standpoint of an external observer of behaviour, the existence of such mental phenomena is something of a hypothesis, for the only thing that "meets the eye" is the occurrence of overt bodily activity (including, of course, linguistic activity). And the mental phenomena which are thus postulated as antecedents of behaviour are, by and large, not identified <u>independently</u> of the actions of which they are the antecedents. Rather their very characterization is expressed in terms of those subsequent actions. Thus suppose I see that you are refinishing the old used dining room table which you have just purchased. It is trivial for me to <u>say</u> that you are doing so because you want to; nevertheless this is what I shall assume. For it would be absurd to suppose that you are doing so though you <u>don't</u> want to. You may, of course, only want to perform the "same" action, as Davidson puts it, "under a different description".[16] Or, as Goldman puts it, it may be an act-token of an act-type which is either "on the same level with"

or generated by the doing of some basic act which is what you want to do.[17] In either case, the very way in which we choose to identify and describe your action should be determined principally by such evidence as pertains to what you, the agent, thought you were doing at the time, whether successful or not. So if it is not under the description which we, the external observers, have originally given it from our external point of view, that an agent can be said to want to do that action, then we would probably revise our description in the light of any evidence suggesting a different description which the action actually had for the agent, the action-description toward which the agent in fact had a "pro-attitude". The application of the two concepts of want, or "pro-attitude", and action is mutually adjusted so that the exact characterizations of the (hypothesized) mental antecedents of action and the overt action are such as to show them to have a kind of "fit" with each other.[18]

The final result is that the application of notions of such mental phenomena as acts of volition, or wants or intentions, implicitly involves a reference to actions which those mental phenomena are supposed to have caused. This point is put, by critics of the causal interpretation of this process, by saying that acts of volition and wants and such bear a logical connection to the acts of behaviour of which they are the purported causal explanation. Because

of this they fail to satisfy the Humean requirement that there can be no internal logical relation between event pairs which are to constitute cause and effect sequences. Because sequences of "volition ⟶ action" or "want ⟶ action" fail to satisfy this Humean requirement of logical distinctness, it is concluded that therefore volitions, wants, and the like cannot be the causes of action.[19] Thus, according to this argument the sort of logical relationships which volitions, wants, intentions, etc. have to action is essentially bound up in the notion of what it is for such mental phenomena to occur at all: this is the notion that the very occurrence of such mental phenomena primarily and intrinsically manifests itself in certain sorts of corresponding overt actions, actions which can therefore logically be expected as the typical manifestation of such mental phenomena. It is thus charged that the mental phenomena are presumably only known to have any of the distinctive characteristics which we know them to have on the basis of our knowledge of their supposed "effects". They are defined in terms of their supposed "effects". It is thus concluded that the culmination of such mental phenomena in the characteristic overt actions in terms of which they are defined, cannot be a contingent matter of fact, nor, therefore, an ordinary causal relationship.

This sort of logical relationship bears a loose connection to that described earlier as constituting the

rationality of action. For if action is understood in terms of its rationality by being understood as described by the conclusion of a practical syllogism the premises of which express the propositional contents of certain wants and beliefs of the agent, then the identity of the want in question at least partly consists in its being a mental condition which, in circumstances corresponding to those that are believed to obtain, ought to lead to the given action.[20] There is the following difference however: I have not expressed wants in so tight a relationship to action that they are nothing more than the wants to perform those actions. Rather, I have conceived wants as more general enduring mental traits of agents, which become achievable now in this situation by action A, now in that situation by action B, and so on as (further) determined by the action-relevant circumstances of those situations.

It is true that if by wanting X, it becomes appropriate for an agent to perform action B in situation Q, then there is a sense in which the agent may be said to want also, in the context of situation Q, to perform action B. This is the sense in which, as Kant puts it, "Whoever wills the end, so far as reason has decisive influence on his action, wills also the indispensably necessary means to it that lie in his power".[21] Thus the sense in which an agent who performs action B also "wants" to perform action B is really more the sense in which an agent simply wills

to perform action B in situation Q, given that this is the necessary means to achieving something (else) that he wants. In such a context, the want which can only be described in terms of the action of which it is a non-contingent antecedent condition is really more the act of choice, or will, or volition which immediately precedes the overt action, than it is a want in the sense of a standing enduring "pro-attitude"[22] of an agent. It seems wise to distinguish between these two sorts of conditions, namely, wants, in a broad sense, and acts of will or volition, particularly in light of recent criticism of the theory of volitions[23] which extends beyond simply the present problem of the logical relation between action and its antecedents. The problem discussed in the present context concerns only the objection raised against a causal interpretation of the human behavioural regularities which postulate acts of volition as intervening between the (standing) wants and the beliefs of the agent and his ensuing action, thus:

Wants + Beliefs ⟶ Volition ⟶ Action

However it is my contention that this objection to volition-type causal theories is mistaken, and that it is not simply because there is a logical relationship between action and its antecedents that a causal interpretation is thereby produced.

Within the context of such causal volition theories,

volitions, or their counterparts, have virtually exclusively the status of, as yet unobserved, theoretical events the occurrence of which is hypothesized as a general account of the occurrence of acts of human behaviour. Because volitions are as yet merely theoretical posits, it is not to be supposed that they could now be described or made intelligible, in particular instances, in any way independently of the particular acts whose occurrences they are specifically designed to explain. Because no independent evidence enables volitions to be characterized in terms of what might be called their "intrinsic properties", they can, at best, be currently conceived only by analogy, in Sellarsian fashion,[24] with the events whose occurrences they (theoretically) explain. This analogical conception of volitions constitutes the only reasonable formal scheme within which the nature of volitions can currently be made intelligible.

In any case, the fact that the conception of particular volitions is currently "logically" dependent upon the conceptions of the particular acts with which they are presumably constantly conjoined, does not entail that particular volitions are themselves conceived as being logically connected with their constant conjuncts. In fact, it is not even clear what a "logical relationship" among actual events might be, unless this was a relationship that happened to obtain among logically relevant contents or

aspects of those actual events, for example, their propositional contents. There is thus a (practical) logical relationship among the intentional aspects of a "want + belief ⟶ action" sequence which, as an actual sequence, happens to instantiate, in respect of those intentional aspects, a practical syllogism. But the causally relevant aspects of that sequence consist in its material embodiments, or, shall we say, its psychological "embodiments". And these aspects neither necessitate, nor preclude the possibility that in other respects such causally necessary sequences may in fact happen to constitute logical sequences.

Thus, there does not appear to be a necessary incompatibility between a logical and a causal relationship among the same phenomena, and hence it is not sound argument to reject the causal interpretation of human behaviour on the grounds of such a view. However this view points toward an important point which can serve as the basis for criticism of those causal theories. This is the point that no exhaustively causal sequence of events as such, without further qualification, is a sequence whose outcome is properly conceived of as being rational. It is only by attributing to each such event an intentional character, a dimension of meaningfulness characterized, for example[6], by means of sentences expressing the appropriate propositions, that the whole sequence, in respect of its intentional

character, may be regarded as rational. But to do this merely as a supplement to a fundamentally causal account of those sequences seems to accord epiphenomenal status to those intentional characters, for they are thereby accorded no causal efficacy in the occurrence of the outcome of the sequence of events (the behaviour itself). This claim shall be defended in the following discussion.

Explicit attention shall be directed toward the mechanistic account of rationality provided in Paul Meehl's "Psychological Determinism and Human Rationality: A Psychologist's Reactions to Professor Karl Popper's 'Of Clouds and Clocks'".[25] Meehl's account happens to be physicalistic in orientation: he argues that the truth of a complete physicalistic micro-causal account of human rational decision-making which contained no explicit reference to reasons, would nevertheless be compatible with a "molar" level account truly asserting the decisive influence of reasons upon the decision. However it is not simply a physicalistic account of human behaviour which I wish to contrast with my form of understanding of human behaviour in terms of its rationality. It is necessary for me to show that it is not simply physicalism but causal determinism in general that is incompatible with the existence of genuine rationality in human behaviour. I shall therefore take the liberty of ignoring the peculiarly neuro-physiological claims of Meehl's argument. His arguments

are to be of interest, then, because they involve the more general claim that a "molar" level account truly asserting the decisive influence of reasons upon decisions is compatible with an exhaustively causal account of the efficacy of those reasons upon the decisions. Meehl's argument is, in essence, that such a causal account is attainable with reference to a micro-physical level of phenomena at which reasons per se do not obtain but instead micro-physical embodiments or correlates of what are called "reasons", with reference to the molar-level of phenomena, do obtain. What is important in the present context, however, is not the distinction between the micro- and molar-level of phenomena, nor the peculiarly physicalistic orientation of Meehl's arguments, but simply the general thesis that an account asserting the genuine rationality of human behaviour is compatible with an exhaustively causal account of that behaviour.

By 'genuinely rational behaviour', I shall mean rational behaviour in the occurrence of which the rationality of that behaviour played an essential role. Thus the logical relationship of the appropriateness of an action, given its antecedents, plays an essential role in the occurrence of genuinely rational behaviour. It would seem that only for behaviour of this sort can the rationality be a matter of anyone's, or anything's, responsibility. For if the rationality of an action, or any event for that matter,

is a merely incidental feature of the process which led to its occurrence, then nothing or no one can be responsible for the rationality of the final stage in which that process culminates. Thus what has to be considered in the present context is the thesis that an exhaustively causal account of human behaviour is compatible with an account of that behaviour as occurring essentially because of its rationality. Put alternatively, the important question is whether or not, if we begin with a causal model for understanding human behaviour, we are justified in supplementing this by a presumption of the genuine rationality of those behavioural sequences.

Let us briefly review Meehl's physicalistic proposal for understanding the causal efficacy of (certain kinds of) mental events.[26] The human brain is to be regarded as being inherently wired in such and such a way and then as being subsequently "programmed" by social learning to have such and such functional connections (dispositions). After having undergone events corresponding to the premises of some argument, such a brain will, with nomological necessity, undergo a brain event corresponding to the logical conclusion of that argument. Thus the defined physical structure of the brain plus the physical laws of brain function causally necessitate that the brain events exemplify, for example syllogistic transitions. It is therefore a physical fact that some particular formal

relation is "physically embodied". The existence of a formal relation of deducibility provides, in a brain for which the theorem obtains, the necessary and sufficient causal condition for a factual transition of inference (a mental process). Meehl acknowledges that, strictly speaking, a reason is not a cause of belief in the conclusion for which that reason is a reason. Meehl clearly has in mind here that a reason in the logical sense of a potential propositional content of some wanting or believing, is not a cause of belief in the relevant conclusion, for he goes on to say that it is the tokening (in the mind of the agent) of the sentence which expresses the reason, that is the real cause.[27]

Meehl's paradigm of rationality is the mental process of simply drawing logical conclusions from given premises. He is less concerned with rationality in the practical sense of being (an action which is) performed so as to achieve wanted ends in circumstances which render it the appropriate manner of seeking to achieve those ends. My account of rationality, on the other hand, is distinctly an account of rational behaviour, and the rationality of behaviour is defined in terms of appropriateness (in given circumstances), to achieving wanted ends. It is true that certain parallels can be seen to obtain between rationality, in a practical sense, and rationality, in a pure, or theoretic, sense. And an account of rationality in either

sense would be strengthened if it could be shown to cover and explicate rationality in the other sense. However in the present context, the discussion of rationality shall be limited to its practical form. Meehl's thesis can be given a corresponding interpretation which is relevant to practical rationality.

One possible difference between theoretical and practical rationality should perhaps be noted. My conception of the rationality of behaviour is built upon a conception of an agent acting in accord with the logic of practical inference. The agent's behaviour is regarded as being performed upon his recognition of the rationality which some possible course of action open to him possesses in a logical sense, that is, in the sense of being representable as the conclusion of a practical inference having as premises statements pertaining to the situation as believed by the agent to obtain and his wanted ends. As I shall argue in the next chapter, I regard such behaviour as the manifestation or culmination of theoretically postulated acts of will or volition. Such acts of will, when they do occur, are conceived as following upon an agent's recognition of the appropriateness, i.e. the rationality, of some course of action regarded as performable by him in the situation which he believes to be at hand. The act of will which is theoretically postulated as initiator (and "causal progenitor", in C. D. Broad's terms[28]) of the

action-culminating sequence is particularly important because it is taken to constitute the vehicle of physical efficacy on the part of the mind with regard to the physical realm. It is only because of the postulation of such theoretical events that the chain of physical causality may be regarded as discontinuous, or interrupted, and the sequences of rational human behaviour can be regarded as contra-causal, or originative. But it is not clear that sequences of pure theoretical reasoning on the part of an agent can be regarded as involving any element which would correspond to an act of will in practical rational sequences.

At any rate, paralleling Meehl's account of theoretical reasoning and the drawing of pure logical inferences, the following account can be constructed for the performance of rational behaviour. Rational behaviour can be construed as simply following upon the obtainment of a set of mental events, or dispositions, which include the wanting of certain ends and the believing that circumstances are such that the given behaviour is rational (in a practical logical sense). In Meehl's account, the sequence consisting of beliefs and conclusion constitutes, because of the logical features of its components, the physical embodiment of what is a formal logical relationship. In the revised Meehlian-style account of rational behaviour, the sequence consisting of the having of a certain want (or wants), the having of certain beliefs about circumstances

(including causal relationships thought to obtain, though these need not be consciously cognized at the time), and the ensuing behaviour constitutes, because of the logical features--that is, the propositional contents--of its components, the mental-physical embodiment of what is also a formal logical relationship. In this case the logic involved simply happens to be that of practical rather than pure, or theoretical, reasoning. Thus, according to such a view, the behaviour physically embodies the conclusion of a practical argument, which can be construed formally as a resolve to act in some specified fashion.

Such an account seems to have the following problems. Firstly, there is a problem which parallels that which was raised as the central difficulty in reconciling the truth of causal determinism with the truth of judgments assigning moral responsibility. This is the problem that because the "person" or the "agent" is not a fundamental category on a causal account, but rather is to be reconstructed in terms of locations for rather coherent unified sequences of events, and because the causal conditions for all these events are ultimately (if causal determinism is true) traceable to sources external to the "person" or "agent", then there is no entity participating in the sequence upon which, or whom, the responsibility for the rationality of action is justifiably pinned. Of course, as Goldman noted,[29] from the agent's point of view, practical

reasoning does not appear as a causal process. The "agent" appears, to himself and to casual observers, as an entity distinguishable from its environment as as somewhat self-contained and "in control" as regards the sequences of events which culminate in its overt behaviour. But although the agent, while deliberating, does not worry about the causes of his action, nevertheless, as Goldman adds, this deliberative process (from the point of view of causal determinism) constitutes a causal process culminating in his action.

The problem for causal determinism, then, is to find some metaphysically significant and fundamental way to distinguish the causal process comprising an agent from those comprising his environment so as to justify the ascription of rational (as well as moral) responsibility to that agent rather than to any external conditions which contributed to his action and rather than to any antecedent conditions which causally generated him as the person he was at the time of action. On the basis of causal determinism, why would any entity ever be responsible for the logical relationship, that is, the rationality, which descriptions of those actions occasionally bore to the positional contents of prior (mental) conditions of that entity. And how could we be so certain that no purely physical entities are ever involved in purely physical processes which could be given similar logical

characterizations in terms of propositional contents, such that those physical entities should be attributed responsibility for the thereby rational events culminating those processes? These are questions which causal determinism appears incapable of answering except in ways which violate commonsense notions about the general human responsibility for rational action. As was found to be the case with regard to ordinary moral notions, causal determinism can be consistently embraced only by disowning certain commonsense conceptions and beliefs. This is an intellectual alternative which I see no compelling reason to take.

Secondly the causal interpretation of rational behaviour fails to illuminate any metaphysically significant respect in which the occurrence of rational behaviour differs from the occurrence of irrational behaviour, that is, behaviour which lacks the proper formal relationship to those believings and wantings of the agent which belong to the set of its causally sufficient antecedent conditions. The formal relationship of rationality, or the lack of it, between the propositional content of the believings and wantings and the behaviour, under the relevant description, is not itself another causal condition within the set of conditions antecedent to behaviour which are causally sufficient for its occurrence. It would seem that, on the basis of causal determinism, the rationality of such

behaviour as is indeed rational in this strictly formal sense is a matter which is thoroughly incidental and epiphenomenal in the occurrence of each particular such rational action.

Thus, on a causal account of human behaviour, it does not seem that the agent can be regarded, in any metaphysical sense, as being responsible for his behaviour, whether rational or irrational. His actions effect or fail to effect his wanted ends in accord with laws of behaviour, alteration of which is beyond human power. If the agent is lucky, then his actions will be rational often enough to ensure for him a good life. We may condemn or applaud him, even punish or reward him, for his actions, depending on the importance which we happen to attach to actions of the sort which he has performed. But we can be justified in doing so only in order to provide conditions which will have a determining effect upon future occurrences of actions of that sort, on the part of the agent in question (and, perhaps, on the part of other agents as well). The agent does not, in any sense, merit the praise or condemnation which we might express, for in no fundamental sense does he merit the responsibility for his actions which we thereby attribute to him. This is not to suggest that the agent is a "victim" of circumstances; such an idea is rendered as illegitimate, in accord with causal determinism, as the contrary idea that an agent "controls" his destiny. The

agent is merely a location for events some of which effect states of affairs which happen incidentally to satisfy the agent by satisfying his wants, and some of which do not. The former happen to be "rational actions"; the latter are "irrational actions".

The sense in which a sequence of beliefs, wants, and action "physically embodies" a practical inference is no more metaphysically significant than the sense in which, for example, mathematical rules might be physically embodied by a sequence of colliding billiard balls which happened to be marked, in the order of their collision, thus: '2', '+', '2', '=', and '4'. It might be true, as Meehl argues, that we could not fully characterize the content of such psychological events as beliefs, wants, and, I take the liberty of adding, actions, without employing the categories of logic. But it does not follow from this, <u>contra</u> Meehl,[30] that the formulation of causal laws of behaviour regarding the influence upon behaviour by what he calls the "tokening" of reasons, that is, believing and wanting, in any way <u>depends</u> upon information regarding this aspect of such psychological and behavioural events. Or, to put it somewhat differently, if the formulation of causal laws of behaviour does <u>depend</u> upon information regarding the categories of logic, or the principles of rationality--and depends upon it <u>in principle</u>, that is, in such a way that only a knowledge of the structure of practical reasoning

and of the correlations between logical categories and physical states of the organism enables us to locate the physical sequences that embody such practical reasonings--then the logical categories, rather than the causal categories, would be primary in our understanding of human behaviour, and would be essential in guiding our discovery of the nomological regularities which involve that behaviour.

It is true that even if causal determinism is true, it may still be heuristically necessary for psychological researchers at this early stage of scientific inquiry into laws of psychology, to continue to rely upon information (presumably drawn from ordinary experience) regarding the logical content of psychological and behavioural events. But unless the logical content of such events in some way figures in the supposed causal process whereby the psychological events exhaustively determine the occurrence of the behavioural events, then the reliance upon such logical information is in principle no more than heuristically necessary. From a causal point of view, such reliance should in principle be eliminable in favour of non-logical causally relevant descriptions of the psychological and behavioural events. For if the logical features of the psychological and behavioural events are simply, as Meehl regards them,[31] entailed nomologically by the features of those events which are the causally relevant features, then the mere fact that such causally relevant features each

entail features which, apart from the causal process, together constitute formal logical patterns and relationships simply cannot, as Meehl would have it, sustain a behavioural level account truly asserting that reasons as such decisively influence the occurrence of behaviour. Some actions, it is true, will be exhaustively determined to occur by sets of antecedent conditions among which will be conditions which, in respect of their propositional contents, constitute good reasons for those actions. But it seems impossible, on a causal account, to avoid construing, as epiphenomenal, and therefore as causally irrelevant, those logical respects in virtue of which events which determine the occurrence of human actions constitute good reasons for those actions.

In his "Of Clouds and Clocks", Karl Popper[32] expresses unease over a similar problem, although for him the important version of determinism, in terms of which rationality remains unexplained and unaccounted for, is what he calls 'physical determinism': the theory according to which the world is described as a physically complete or physically closed system proceeding in accordance with its own laws of interaction which leave no room for interference by anything outside that system. For Popper, "It is this 'closure' of the system that creates the deterministic nightmare".[33] According to Popper, physical determinism is inadequate because it is unable to account for such obvious

everyday facts as the effect which our behaviour has upon the physical world in virtue of what Popper calls the "abstract and non-physical entities" such as reasons, arguments, and rules that appear to have so fundamental a status within the mental realm--a realm which, furthermore, appears to determine how we behave.[34]

I have been suggesting that the version of determinism inimical to the notion of rationality and free will is that of causal determinism whether it is conjoined with a strict materialism or is instead conceived as applying also to an additional mental realm. As I argued in chapter four, causal determinism of either sort fails to generate a significant metaphysical basis for distinguishing the behavioural and mental processes of a human agent from the processes of the environment external to the agent, both contemporary with and antecedent to the behavioural and mental processes in question. The problems pertaining to rationality and responsibility, for an exhaustively causal metaphysical view, are just as serious whether behaviour which is incidentally rational, is conceived as occurring through the causal efficacy of material, or of psychological, conditions--or of both together.

It is difficult to believe that the pervasive rationality of human behaviour could be a mere incidental by-product of the causally irrelevant propositional contents of otherwise only causally necessitated mental and

behavioural regularities; yet causal determinism in itself provides no more enlightening an account of rational behaviour than this. What we seem to feel in ordinary contexts is that "abstract and non-physical entities" such as are listed by Popper:[35] namely, purposes, deliberations, plans, decisions, theories, intentions, and values, play a fundamental part in the behaviour which we perform and the ensuing changes in the physical world. Causal determinism is simply unable in itself to account for such apparently obvious facts.

Popper draws a useful distinction between two different problems that are at issue here. One is what he calls "Descartes' problem": the familiar classical problem of how such mental phenomena as volitions can influence or control the physical movements of our limbs.[36] It should be noted, in passing, that this problem remains as much a problem for non-exhaustive materialistic causal determinism as it is for any non-causal account. The second problem to which Popper calls attention is one which he describes as "the problem of the influence of the universe of abstract meanings upon human behaviour (and thereby upon the physical universe)".[37] The "abstract meanings" in question are, again, such things as promises, aims, rules--more specifically the propositional content or meaning of such physical phenomena as documents and pronouncements,[38] and also of such psychological phenomena as thoughts. The "universe of

abstract meanings" is what Popper, in another place,[39] calls ". . . the world of intelligibles, or of ideas in the objective sense; it is the world of possible objects of thought: the world of theories in themselves, and their logical relations; of arguments in themselves; and of problem situations in themselves". Finally, borrowing from Frege, Popper calls it the world of "objective thought contents".[40] The problem to which I have attempted to draw attention pertains also to this "universe of abstract meanings". It may be expressed as the problem that causal determinism contradicts apparently commonsense beliefs by denying the real influence of practical reasoning, in the sense of the logical relationships among "objective thought contents", and the practical rationality of action upon the world. Instead these "abstract meanings" are relegated, in whatever form they are capable of obtaining compatibly with the truth of causal determinism, to a purely epiphenomenal status. From the viewpoint of commonsense, this conclusion as it stands is an unacceptable consequence of causal determinism.

Suppose that formal patterns of rationality which are exhibited by "want + belief ⟶ action" sequences can be expressed as rules enjoining the performance of those actions given the want of certain ends and the existence of certain circumstances as would be characterized by the statement expressing the appropriate propositional contents

of the relevant beliefs, for example, "If you want to maintain your car in good condition, then at approximately every three thousand miles you should change the oil". Let us call such statements "rules of rationality", or "principles of action". My argument might then be expressed by saying that in accord with a causal account of human behaviour, actions can do no more than merely conform to rules of rationality. I am contrasting "mere conformance to rules" with something a good deal more than this, in accord with which it would be justifiable to say that an act occurred because, or in virtue of, the rule of rationality to which it (also) conforms. What this "good deal more" seems to be is a sense or appreciation, on the part of the agent, that the action which he's performing is right, appropriate. This does not require that an agent consciously apprehend as such the principle of action in accordance with which his action is appropriate. Such a requirement would certainly exceed the usual circumstances of human action. But if a human being is conceived as an essentially rational being, then this is to conceive him as a being who is essentially capable of sensing and appreciating the rightness, appropriateness of a projected course of action in circumstances as he believes them to obtain. To have such an appreciation, and to act accordingly given such an appreciation, is what I mean by acting in virtue of the rules of rationality, or principles of action. This view

of how human action comes to be performed is fully consonant with the commonsense assignment of responsibility to an agent for the performance of an action, indeed, is the very least which seems requisite to sustaining the justification of such assignments. Because action occurring in accordance with causal determinism would merely conform to rules of rationality, then, in accordance with causal determinism, the basis for holding persons responsible for their behaviour is thereby eliminated, a conclusion which is unacceptable from the point of view of commonsense.

Sellars' account can be extended to cover not only the linguistic rules with regard to which it was explicitly formulated, but "rules of rationality" as well. In accord with such an extended account, the performance of a certain type of behaviour which happens to be appropriate in a given type of situation to satisfy a certain sort of want of the organism, will come to be reinforced as a repeatable form of behaviour (given appropriate circumstances) in the repertoire of the organism as a direct consequence of the survival value which such rational behaviour has for the organism, that is, as a direct consequence of enabling the organism to have its wants satisfied and its goals achieved. The explanation of human actions would thus parallel the form of explanation given of evolutionary changes, in biology. Evolutionary changes are taken to occur because of the biological rapport which they make possible between

species and environment. According to Sellars, this sort of explanation is to be construed as a statement concerning the consequences to particular organisms, and hence tc their hereditary lines, of standing or not standing in relations of rapport with their environment. The explanation, Sellars warns, is not to be construed as one which, in Sellars' words, attributes "causal force to an abstraction", consequently tempting us to ". . . introduce a mind or minds to envisage the abstraction and be the vehicle of its causality. . . ."[42]

However this extended Sellarsian account of behaviour as being "pattern-governed" by "rules of rationality" still does not seem to render the causal account of behaviour consistent with the thesis that actions which are rational occur essentially in virtue of the rules of rationality to which they conform. Firstly, the account would only enable us to explain, "in terms of" its rationality, behaviour which in fact contributes to the survival and betterment, in a biological sense, of individuals of a species. Yet the formal relationship of rationality is one which may obtain in regard to behaviour which has no survival value whatsoever because it is behaviour which contributes to the satisfaction of desires and purposes on the part of the organism which are either neutral with regard to, or positively detrimental to, the survival and genuine biological betterment of the organism. Such behaviour may not

be rational in terms of what will contribute to the particular "end" of long-run betterment of a species; but it is nevertheless still rational in the individualistic sense which I have been employing, namely, in the sense of constituting the means to achieving the ends wanted by an individual. In this "individualistic" sense of rationality, the desirability of wants qua wants is never open to question. Wants are simply taken as given for the purpose of understanding the behaviour which has been performed in order to achieve their satisfaction.

This account might be revised in the following fashion to take account of the above objection: organisms which, in general, act rationally in terms of their given wants and beliefs--whether or not these are useful and adequate reflections of their objective conditions in the world--will be more favoured in terms of survival than will organisms which do not. Though this principle is not unplausible as it stands, it is hard to see how its rather intuitive plausibility can be empirically cashed in. Have not all plants and a vast number of animals survived, perhaps even thrived, without exhibiting anything that we could call "genuinely rational behaviour"? It is true that on Popper's "problem-solving" account of the behaviour of all living organisms, even the behaviour of plants is to be understood as the plastically-controlled response to an objective problem-situation,[43] and in this sense as rational.

However unless conscious awareness of a problem situation as a problem-situation can, in principle, be pointed to as a condition which shows the situation to be really "a problem" for the organism, then there seems to be no reason why we should differentiate the activities of organisms from purely material non-living processes of the universe. All finite causal sequences merely end somehow without being necessarily rational thereby. If we cannot suppose that an entity wants or intends the end-state of a sequence in which it is involved, then it is not clear why the sequence should be reconstructed as a process of activity aiming at this end-state as the solution of a problem, rather than as an ordinary causal process in which "aims" play no part. And if we suppose that the aim at this end-state does somehow participate in the process whereby the organism's activity comes about, then we do seem to be on the slippery slope which will lead us to attribute "causal force to an abstraction" and consequently tempt us ". . . to introduce a mind or minds to envisage the abstraction and be the vehicle of its causality. . . ." And where the end in question is that of an organism which is itself non-conscious, then the slope is slippery indeed.

Secondly, what an evolutionary theory of this sort appears to be able to explain is the "operation" of rules of rationality within the behavioural repertoire of a species. It is difficult to see how an extended Sellarsian

account of rules of rationality would enable us to explain individual behaviour in terms of its rationality as the behaviour of an individual. The difficulty may best be brought out by considering some of Sellars' remarks on the application of this manner of understanding behaviour to the phenomenon of the dance of the bees. Sellars claims that the first behavioural exemplifications of such dance patterns by particular bees is not behaviour which is appropriately described by saying that successive acts occur because of the pattern. Presumably, before the forces of evolutionary reinforcement and extinction have been able to affect the bees' behaviour, the formal patterns exhibited by the behaviour in question must indeed be thought of as merely rule-conforming behaviour. It is only after natural selection has operated to ensure that all bees have the internal "wiring-diagram" that neurophysiologically accounts for the behaviour as a particular kind of rule-conforming response to environmental conditions, that it is appropriate to describe the particular acts of the dance behaviour of the bees as being actions which occur because, or in virtue of, their role in the dance pattern as a whole.[44]

Yet, at an individual level, the behaviour of those bees which "danced" in olden times is (presumably) in no respects different from the behaviour of individual bees who "dance" in contemporary times, except perhaps in respect of

the social non-conformity which the ancient bees represented within the total bee population of their day. If it was indeed the internal "wiring diagram" which accounted for dance behaviour on the part of individual bees, then it would seem that the same factor should account for the dance behaviour on the part of individual bees now. Natural selection accounts not for individual behaviour as such, but rather for the pervasiveness of behavioural patterns throughout a species.

While these remarks may show that an evolutionary account will be of no help in constructing a causal account of individual human behaviour, they do not yet show that a causal account cannot be constructed, compatibly with commonsense explanations of human behaviour in terms of its rationality. It is true that we do not need to suppose that for any given event there is only one single unique explanation that constitutes "the explanation" of the event, adequate to any and all explanatory contexts. Thus to take an example,[45] suppose that we are attempting to explain why an adding machine has printed the number 101,187 on a tape after the numbers 70,445 and 30,742 have been punched into it by means of its keys. Clearly there is a mechanical account of the operations of the adding machine which can be given in terms of gears and wheels and printing keys and significant marks on paper, an account which is perfectly intelligible and correct of its kind. But is there not also

another intelligible and correct explanatory account which can be given for the activities of the adding machine? To a question of why the adding machine printed 101,187 on a tape, can the response not consist of simply citing the facts that the numbers 70,445 and 30,742 were punched into it, and that 70,445 and 30,742 equal 101,187? It would be tacitly assumed in the giving of such an answer that the adding machine operates properly, and this means that it is precisely the sort of mechanical contrivance whose internal gearing and related mechanisms are such as to result in printed symbols which represent the proper computations based on the "raw data" introduced into the machine. That is, to explain the activities of the adding machine in terms of raw input data is to tacitly assume that it is the type of entity which constitutes a (physical) embodiment of certain rules for handling that data, namely, the laws of arithmetic. In the context of such an assumption, the explanation by reference to significant input data seems perfectly intelligible and correct of its kind. The explanation of the adding machine's activities in terms of internal gears and what not may be said to constitute an explanation of how the physical system comes to <u>constitute</u> a physical embodiment of the rules of arithmetic, and this sort of explanation clearly operates at a different level than the explanation in terms of input, for this sort of explanation is more nearly an explanation of the background

assumption in the context of which the explanation in terms of input becomes possible. It follows that the two sorts of explanatory account, rather than competing with each other and being mutually exclusive, are quite consistent with, and even supplementary to, each other. Notice that only the mechanical account is a causal account. Yet the correctness of this account does not undermine the correctness of an explanation of the adding machine behaviour in terms of arithmetical rules.

The example of an adding machine can be regarded as a material analogy for the "psychological system" of a human being which psychologically "embodies" certain rules, in this context, the rules of rationality, or principles of action, in particular. Based on the analogy, it might be argued that a causal explanation of rational behaviour is by no means necessarily incompatible with an explanation of behaviour in terms of such "input data" as reveal the rationality of the behaviour in question. An explanation of behaviour in terms of such "input data" as reveal the rationality of the behaviour in question remains precisely an explanation in terms of <u>abstract entities</u>: intentional features, "objective thought contents"--in short, the propositional contents of the wantings and believings of the agent. The explanatory relevance of these factors is preserved at the level of understanding human behaviour in terms of its rationality, yet the entire account <u>seems</u> quite

consistent with a causal account.

But still certain problems remain. We are far from holding adding machines responsible for the arithmetically accurate "responses" which they exhibit to given input data. Furthermore we must consider how the activity of an adding machine comes to represent the embodiment of, i.e. materially instantiate, arithmetical rules. That is, over and above the causal account of the machine in terms of internal gearing mechanisms and so forth, we can ask how the activity and output of the adding machine come to mean arithmetical computations, given that the machine is, after all, a purely physical system. The immediate answer is that adding machines are the contrivances of human beings, and are contrived intentionally so as to constitute just such material embodiments of laws of arithmetic as we find them to be. Human beings build adding machines precisely by arranging the internal gearing mechanisms and what not so that these will result in the orderly print-out of numerical signs in a way that is arithmetically accurate given the numerical interpretation that we are to put on the keys initially punched. We thus know that the adding machine constitutes a material embodiment of the laws of arithmetic because we know that it is contrived to be so.

But no similar analogy may be found in the case of a human being, or a human mind, unless we suppose some further "being" who analogously contrived it to be that

human beings embody the rules of rationality. A mere physical system, untampered with by human beings, does not, in itself, "embody" rules of any sort, in the sense of constituting a material interpretation for a structure of generalized formulae that has significance independently of the functioning of that system. A mere physical system, let us suppose, is causally determined to proceed through successive changes in accord with, that is, in mere conformance with, the relevant causal laws which describe the behaviour of that system. Those relevant laws are formulated precisely with a view to expressing the regularities involved in the successive changes of the system; the empirical significance of the laws consists in the very regularities of the system, and others like it, which proceed in accord with the laws. The system itself is given in the world prior to our formulation of relevant and applicable laws; our formulation of these laws may be said to constitute an intellectual "embodiment" (note: not a psychological "embodiment") in the realm of ideas and "objective thought contents" of the regularities which we discern in our experience of such a system.

The relation between an adding machine and the laws of arithmetic is one in which this order is reversed. The idea of the system of laws or rules is given first, in our intellectual apprehension of it, and the material system is organized so as to conform to this system of laws or rules.

Its conformance is, in this way, not mere incidental conformance to laws but intended conformance to rules. The activities of the material system proceed as they do because of the rules of arithmetic, and they do this because they have been arranged that way precisely by a process in which an organism capable of "envisaging abstractions" such as rules of arithmetic could be the "vehicle" of the "causal force" of this system of abstractions.

It is true that a non-contrived physical system can possibly be regarded as the material embodiment of a system of rules even if no human being arranged that system to operate as it does. That is, a natural or "found" physical system may be amenable to interpretation in accordance with a system of rules or laws that is not merely descriptive of the mechanical interactions obtaining within the system. Its amenability to such interpretation would have to be thus given in the system itself and would therefore not be introduced by contrivance. Nevertheless, a "found" physical system can only be thus regarded as a material embodiment for a system of rules if various stages in its natural operations can be interpreted meaningfully or significantly so that the (artificial) causally generated patterns which emerge among the meanings correspond systematically to a set of rule-governed interrelationships among those meanings as such.

Neither an adding machine nor any other physical

system gives meaning to its own physical states, nor uniquely indicates "the interpretation" which is to be put upon its activities--or, even, that any interpretation at all is "to be put" upon its activities. The arithmetical significance of the activities of an adding machine is something which we assign to it; we decide upon such an interpretation in accord with our conception of the system of arithmetical rules. The interpretation is "for us"; it is not something for, of, or by the physical system in itself. This is as true of natural physical systems as it is of contrived machines or tools. Thus, in order that there can be the physical embodiment of a set of rules, there must be an external interpreter who can correlate meaningful abstractions with the physical processes exhibited by the system.

If human beings are really psychological or psycho-material embodiments of rules of rationality, then it seems that we could legitimately raise the question of how the succession of causally necessary human states come to be interpreted meaningfully in accord with those rules. Some sort of theistic account is possible. According to such an account, meaning would be assigned extrinsically as it were, to various stages in the natural causal sequences of already orderly human conditions. However such an account would seem to have the defect that only divine revelation would enable us, the human beings in question, to know the sort of

interpretation in accord with the rules of rationality which is necessary to understanding human behavioural sequences in terms of those rules. Yet it hardly seems that so elaborate a procedure is required for us to reach the understanding that we do in fact achieve of human behaviour--not only that of other human beings of whom we are the "external observers", but as well that of our own behaviour.

A working hypothesis, which is not altogether implausible, is that the meaningfulness of human behavioural regularities is not simply assigned externally, but rather, that human beings themselves, regarded as a species, seem to imbue their own states with meaning and significance. This would suggest that the primary, though not necessarily infallible, source of information regarding the rational meaning and significance of a human being's condition and activities is to be regarded as the report of that human being himself. Human beings have states of conscious awareness which appear to be intrinsically "of" other things, and in that way, apparently intrinsically intentional or meaningful for the very organism undergoing those states. This does not seem to be something which can reasonably be said of physical systems. The significance of physical systems appears to be ultimately a matter of interpretation. On the other hand, in accord with the working hypothesis above, human beings would be regarded as self-meaningful;

they themselves would be seen as the source and origin of the very meaningfulness which permits their activities to conform to the rules of rationality (over and above the presumed conformance of those activities to causal laws).

We seem to be led to the strange conclusion that human beings interpret their own natural causal activities in accordance with rules of rationality. On a causal interpretation of such activities such a conclusion is ridiculous, for human beings do not exist independently of their own natural causal activities or have an independent standpoint from which to apply such interpretations to "themselves". This is the point at which the adding machine analogy becomes useless, for at this level the two cases are no longer analogous. There does not seem to be any way to reconcile the causal account with a theory of self-meaningful behaviour, which is what rational activity appears to have to amount to in the case of human beings.

Can this notion even be made sense of at all? Some useful suggestions can be derived from the discussion, at the beginning of the present chapter, concerning the concept of "reasons" as constituting a relative conception: relative, that is, to the functioning of such entities in wider sequences which constituted processes of reasoning. Thus it could be said that the intrinsic, or self-, meaningfulness of human mental states is intrinsically bound up with the role of those states in orderly mental

(and behavioural) processes. That is, the meaningfulness and the orderliness of human mental states cannot be separately accounted for; the account of either feature must essentially involve an account of the other feature. (We understand what sort of "reason" a particular want is by understanding the sorts of actions which it rationalizes in different situations, that is, by understanding its practical implications.) This point is, in effect, suggested by Popper in "Of Clouds and Clocks" when he says that the ". . . controlling power of meanings such as the contents of our theories, or of purposes, or aims . . ." is ". . . part and parcel of these contents and meanings; for part of the function of contents and meanings is to control".[46] The important "control" which such contents and meanings exert upon us, from the point of view of the present context, is the imperatival force which rules of rationality, expressing what ought to be done in given circumstances to achieve wanted ends, have upon beings who will the indispensably necessary means to the attainment of their ends. If this account of meaning and rational order is correct, then by being "self-meaningful" in their mental and behavioural states, human beings are also at once being "self-rational", for an intrinsic part of the meaningfulness of practical rational states and events consists in their (rational) ordering.

A direction thus seems to be vaguely discernible

in which we can go to elaborate the psychological means by which a human being constitutes the psychological (and behavioural) "embodiment" of rules of rationality. However if that direction has been at all correctly indicated, then by following it we shall be taken far afield from the causal theories of rational behaviour with which we began. We seem to be able to regard human beings as the psychological "embodiments" of rules of rationality only by regarding them as, in effect, the original sources of their own ordered and meaningful activity. Rather than being able to divide the account of meaningfulness--an essentially rule-guided account--from the account of orderliness--by itself, an essentially causal account--as we were able to do in the case of adding machines, and therefore find the two accounts compatible with respect to the same system, in the case of human beings, we seem to have to conflate them. Hence by supposing that the meaningfulness of human states and activities is self-original, we concurrently come to regard the orderliness of human states and activities as self-original, and thereby exclude the possibility of mere causality as an exhaustive account of regularities at that level. To put this same point in a more methodologically-relevant fashion, the explanatory framework for understanding human behaviour that is oriented around the notion of rational action as the (free) self-original activity of human beings is basically a framework which is to be

applicable to a kind of system the inherent nature of which determines behaviour in accord with a <u>form</u> of laws not found in purely physical systems and not derivable from the laws of physics and chemistry which pertain to the properties and activities of the sort of (physical) constituents as comprise the complex human wholes, as these constituents occur separately from complex human systems.

FOOTNOTES

[1]Some proponents of this sort of interpretation of behavioural regularities are the following: William Alston, "Wants, Actions, and Causal Explanation", in Hector-Neri Castaneda, ed., Intentionality, Minds, and Perception (Detroit, 1967), pp. 301-341; Bernard Berofsky, "Purposive Action", American Philosophical Quarterly, vol. 7 (1970), pp. 311-320; Donald Davidson, "Actions, Reasons, and Causes", The Journal of Philosophy, vol. LX (1963), pp. 685-700; W. D. Gean, "Reasons and Causes", Review of Metaphysics, vol. 19 (1966), pp. 667-688; Alvin Goldman, A Theory of Human Action (Englewood Cliffs, 1970), esp. ch's. 3, 4, 5, 6; Charles Landesman, "The New Dualism in the Philosophy of Mind", Review of Metaphysics, vol. 19 (1965), pp. 329-345; Ruth Macklin, "Reasons vs. Causes in Explanation of Action", Philosophy and Phenomenological Research, vol. XXXIII (1972), pp. 78-89; Ruth Macklin, "Action, Causality, and Teleology", British Journal for the Philosophy of Science, vol. 19 (1969), pp. 301-316; Paul Meehl, "Psychological Determinism and Human Rationality: A Psychologist's Reactions to Professor Karl Popper's 'Of Clouds and Clocks'", in Michael Radner and Stephen Winokur, eds., Minnesota Studies in the Philosophy of Science, vol. IV (Minneapolis, 1970), pp. 310-372; Stephen Toulmin, "Reasons and Causes", in Robert Borger and Frank Cioffi, eds., Explanation in the Behavioural Sciences (Cambridge, 1970), pp. 1-26; and Morton White, Foundations of Historical Knowledge (New York, 1965), esp. ch. 5. Carl Hempel belongs on this list as well, but his work shall be given special consideration in the next chapter.

[2]For present purposes, it is irrelevant whether or not those "mentalistic" phenomena are ultimately to be given a materialistic account. The causal thesis can be formulated separately from a materialistic thesis that so-called "mentalistic" phenomena exist only in some manner of material embodiment, or are efficacious only in that form.

[3]Reprinted in May Brodbeck, ed., Readings in the Philosophy of the Social Sciences (New York, 1968), p. 50.

[4]A Theory of Human Action, p. 78.

[5](Oxford, 1972), especially "On the Theory of the Objective Mind", pp. 153-190, and "Of Clouds and Clocks", pp. 206-255.

[6]Cf. "Purpose, Function, Scientific Explanation", in Brodbeck, ed., op. cit., pp. 215-216.

[7]"Psychological Determinism and Human Rationality: A Psychologist's Reactions to Professor Karl Popper's 'Of Clouds and Clocks'", p. 317.

[8]Ibid., pp. 317-319.

[9]This point is made by W. D. Gean, "Reasons and Causes", p. 672.

[10]Op. cit., pp. 46-47.

[11]Ibid., p. 46.

[12]"Reasons and Causes", pp. 19-20.

[13]"The New Dualism in the Philosophy of Mind", p. 341.

[14]Op. cit., pp. 112-113.

[15]Ibid., p. 115.

[16]Op. cit., p. 46.

[17]Op. cit., pp. 54-55.

[18]As Charles Taylor puts it, an essential element in the classification of action as an action of a certain type, i.e., as directed to a certain goal, is the goal to which it is directed by the agent, and therefore the intentional description which the action has for the agent qua action by the agent. The Explanation of Behaviour (London, 1964), p. 58.

[19]For examples of this argument, see: A. I. Melden, Free Action (London, 1961), ch. 8; Richard Taylor, Action and Purpose (Prentice-Hall, 1966), pp. 52, 68; Charles Taylor, op. cit., p. 44.

[20]Not in a moral sense, but in the practical sense that one ought to undertake the means necessary to one's given ends.

[21]Foundations of the Metaphysics of Morals, trans. by Lewis White Beck (Liberal Arts, 1959), p. 34.

[22]Davidson's term. Cf. op. cit., p. 46.

[23]Cf. Gilbert Ryle, The Concept of Mind (New York, 1949), esp. ch. 3.

[24]Cf. Wilfrid Sellars, "Empiricism and the Philosophy of Mind", Science, Perception, and Reality (London, 1963), pp. 186-195; and Science and Metaphysics (London, 1968), pp. 18-26.

[25]Op. cit.

[26]Ibid., pp. 321-323.

[27]Ibid., p. 329.

[28]Cf. "Determinism, Indeterminism, and Libertarianism", in Bernard Berofsky, ed., Free Will and Determinism (New York, 1966), p. 152.

[29]*Op. cit.*, p. 80.

[30]*Op. cit.*, p. 317.

[31]*Ibid.*, pp. 356-357.

[32]In: *Objective Knowledge: An Evolutionary Approach*, pp. 206-255.

[33]*Ibid.*, pp. 218-219.

[34]*Ibid.*, p. 225.

[35]*Ibid.*, p. 229.

[36]*Ibid.*, p. 231.

[37]*Ibid.*, p. 230.

[38]*Ibid.*

[39]"On the Theory of the Objective Mind", also in *Objective Knowledge: An Evolutionary Approach*, pp. 154-155.

[40]*Ibid.*, p. 156.

[41]In: *Science, Perception, and Reality* (London, 1963), pp. 321-358, esp. pp. 325-327.

[42]*Ibid.*, p. 326.

[43]Cf. "Of Clouds and Clocks", *Objective Knowledge: An Evolutionary Approach*, pp. 206-255.

[44]Op. cit., pp. 326-327.

[45]The following way of putting the point was suggested to me by Professor Robert W. Binkley.

[46]Op. cit., p. 240.

VI

THE THEORY OF AGENCY

As argued in the preceding chapter, the foundation of a non-causal account of rationality might consist in the premise that human beings are self-meaningful and self-rational. This almost seems to be a claim that human beings exist as rational somehow prior to and independently of their own process of rational activity, which must then constitute the rational orderings into which they originally contrive themselves to think and behave. Such a claim might make sense if we were able to distinguish a human being in two respects: one, as a "person", the public mental-physical entity that interacts with people and has character, wants, beliefs, etc.; and two, as a "self" or "agent", the as-yet undefined origin of the self-meaningful and self-rational activity characteristic of human beings.

What would be grounds for thinking that anything plausible and substantial could be made of such an account? Firstly, the previous discussion of the notion of responsibility is the source of one favourable consideration. For it seemed that a condition for the application of the notion of responsibility was the requirement that there be an "agent" distinguishable from all the external and

internal processes that were accountable on purely causal grounds. It seemed that only to an agent capable of contra-causal originality, hence one which constituted an origin of activity, could responsibility for action justifiably be ascribed. The "agency" account of rationality is preferable, in this respect, to the causal account.

Secondly, many authors who embrace causal interpretations of human behaviour appear to do so in the first place only because they fail to conceive any viable alternative view. Thus, as noted earlier, a problem for the explanation of behaviour in terms of its rationality seems to be that those abstract entities which rationalize and justify behaviour do not appear to have explanatory force at a more enlightened level of explanation. The logical relationship of rational justification does not <u>in itself</u> seem to constitute a sufficient reason to expect the occurrence of that which is rationally justified even in terms of <u>existing conditions</u>. Generally, the view is accepted that there is <u>some</u> relationship between wants and beliefs which precede action, and ensuing action; the issue is to find a satisfactory and compelling account of just what <u>sort</u> of relationship this could be. Thus, characterizing Aristotle's attempt to explicate this relationship as one in which the concept of wanting is introduced as a <u>causal</u> factor, Davidson remarks that "Failing a satisfactory alternative, the best argument for a scheme like Aristotle's

is that it alone promises to give an account of the 'mysterious connection' between reasons and actions".[1]

If, however, an alternative theory can be worked out along the lines which I have suggested above, an alternative which I shall call an "agency" theory of rational behaviour (and thus bear out the concluding remark of chapter four that in my view a metaphysics of rationality dovetails with a metaphysics of agency) then we do not need to embrace a causal account of rational behaviour <u>just</u> because it alone "promises to give an account" of the relation involved in the "want + belief —— action" sequences. I suggested earlier that the causal theory was resorted to because it could not be conceived how logical relations could otherwise be efficacious in the production of events except through the embodiment of those relations in conditions which could produce further events through causal efficacy. The agency theory of rationality provides us with the means for making sense of an alternative way in which logical relations, indeed Popper's entire universe of abstract meanings, could be efficacious in the production of events. The sort of account which the agency theory enables us to formulate is the following: an agent is an entity capable of <u>acting upon his appreciation</u> of the rationality of a course of action open to him, as determined in accord with his wants and the circumstances as he believes them to obtain; and a human agent is, furthermore,

a being intrinsically capable of rational apprehension and appreciation. In this way the logical relationships between wants, beliefs, and the actions which they rationalize constitute essential explanatory factors pertaining to behaviour, because we are precisely precluded by this account from explaining behaviour in terms of the agent's apprehension of abstract meanings which render a course of action open to him either irrational or rationally irrelevant. In accord with the fundamentals of this view, such reference is simply unacceptable as explanation of anything which is to be regarded as distinctively human behaviour (as contrasted with the merely physical changes of those beings).

I have suggested that causal theorists may be resorting to causal accounts of rational behaviour for want of a viable alternative. It is interesting to note that the presentation of such causal accounts is occasionally accompanied by apologetic qualifications which may serve implicitly to smuggle in some non-causal notions which lend the causal accounts an air of greater intuitive plausibility yet which are more in the spirit of the creative agency theory elaborated here. For example, Goldman suggests that not all causation need be assimilated to mechanical causation, that causality need not be taken to imply blind, unreasoning mechanism.[2] This qualification is meant to allow for the possibility that intentionalistic states, and,

hence, reason, can be involved in causal relations. But if my previous arguments have been correct, then the pursuit of this line of thought involves ultimate abandonment of the idea that those "relations" can be merely causal. Stephen Toulmin claims[3] that even though causal mechanisms are present in all cases of behaviour, we can still distinguish actions "done for reasons" as being actions which involve the use of procedures (of calculation, etc.) that we have learned during our lifetimes, from actions which by-pass all deliberation, e.g. actions caused by brain stimulation. Toulmin suggests that the "compulsion" present in the deliberative cases would merely be the "compulsion" of compelling reasons, a compulsion which in no way impugns our responsibility. The reasons for which a man acted are the considerations which carried weight with him in his deliberations because he has learned to recognize their relevance to his deliberations. Any internal neurophysiological basis for such actions, if considered in isolation from the "weight" which different external considerations carry for us, fails to make fully intelligible the actions which we do "for reasons". Toulmin therefore suggests that the actions performed "for reasons" be regarded as subject to a causality of their own.[4] But a causality "all its own" may not be a causality except in name. There seems to be no need to retain the use of a label the original understanding of which involved the use of a certain model if we do not

also retain the use of this interpretive model. The model for causal sequences seems to be the model of a conservative system in which the antecedent conditions which "determine" subsequent events are themselves exhaustively determined by conditions external to the agent.

Brand Blanshard goes even further than other causal theorists. He agrees that if determinism involves either materialism or mechanism, it would reduce to absurdity any notion of moral responsibility. Given this, he suggests that in the realm of reflection and choice there operates a kind of causality which is different from any that we know in the realm of bodies. In the psychological realm there are indeed several different levels of causality, from the simple psychological law of association all the way to laws of logical inference in which, as Blanshard puts it, thought is determined in its progression by the necessities in its object.[5] Blanshard is arguing particularly against indeterministic accounts of moral responsibility which focus on cases where one sees that one ought to do y, and does it, despite habit, impulse, and association prompting one powerfully to do x. Blanshard regards action impelled by a sense of duty as essentially reasoned action, that is, as action which is determined by reason. Thus, rather than being undetermined, such action is determined by the moral necessities of the case, which is the moral man's freedom. In alternative terms, such

action is rationally determined, and for Blanshard who espouses a plurality of kinds, or levels, of causal determination, rational determination is of the highest level.[6] Again the suggestion can be made that the use of the term 'causality' is misleading in this context if it is not to be associated with such notions as that of merely conservative transmissive systems. What Blanshard characterizes as the higher levels of causality seem to exceed the capacities of the conservative model, and should perhaps be regarded in the light of an entirely different metaphysical model: namely, the model of the creative agent.

The agency theory of rationality which I wish to support[7] may be briefly characterized in the following way. I have been contending that free action (which I take to be a genuine and intelligible possibility in the world) is action the choice or will to perform which has not been exhaustively caused to occur by conditions external to the performing agent. I suggested (in chapter four) that it is the distinctive role of reason and rational activity which can become the basis of a genuine understanding of freedom of the will, or mind. (As B. A. O. Williams has noted in a similar vein, ". . . there is more to freedom than freedom of will".[8]) I have suggested that rational action is action which may be said to occur for the sake of a goal or state of affairs which is expected to obtain as a result of the performance of the action. It is action which can be taken

to correspond to the conclusion of a practical syllogism the premises of which can be taken to correspond to the beliefs and wants that are among the antecedent determining conditions of the action. However, and this point is essential, it does not follow from the mere fact of such a determining, or nomological, relationship that the antecedent conditions of the action are also its causal conditions. Genuinely rational action, I have suggested, is action the determination of the occurrence of which somehow incorporates its rationality, or justification, as defined above, among its antecedent conditions. Consequently rationality, or justification, is an explanatorily relevant factor in regard to genuinely rational action. I have tried to show that such action, as a type of event for which the behaving entity can be held "responsible", cannot be understood by the use of the concept of causality as a model for the sort of determination which is involved, that genuinely rational action requires contra-causal conditions. It is in this way that the concept of genuinely rational action seems to dovetail with the concept of free action, a point which is reinforced by certain ordinary applications of moral notions. To be a matter of responsible behaviour, such action must be understood in terms of a real origin of activity. This notion may plausibly be made sense of, given the attendant notions just outlined, by supposing rational action to be

performed upon the appreciation of its rationality by a performing agent, a being essentially rational and essentially the real original source of its own activity in accord with the rules of rationality, a being essentially "self-legislative" and "self-rational".

I have also been maintaining that free, rational action can be conceived as determined action, that it is not to be thought of as requiring indeterminism simpliciter for its possibility, but only a measure of causal indeterminism. There is another kind of determination which may be conceived to obtain in the relationship between reasons, agents, and actions, and which is already implicit in the notion of an inherently rational being. This alternative kind of determination may be called "rational determination". Action which is rationally determined to occur is precisely an event the rationality of which must somehow be involved essentially among its determining conditions. Rational determination is thus a concept inseparably bound up with that of genuinely rational action, and is the manner of its coming about. The notion of an inherently rational being is the notion of a being who is determined to act upon his appreciation of the force of reasons, a being who is thus rationally determined in his activities. Because he is self-rational, this determination is a matter of original activity on his part and not the causal consequence of mental phenomena which happen to

constitute "reasons" for him.

What I am calling rational determination is that which has also been known under the term 'final causality'. The wanted goal or end-state may be said to (logically) determine its means, that is, the action which is rational in given circumstances, just by virtue of the fact that that action is appropriate for achieving the specified goal or end-state. But the want of such an end does not causally necessitate the action which it rationalizes. It is for an inherently rational, active being to act upon, in accord with and in virtue of, his recognition of the rationality of an action, that is, his recognition of what he ought to do in existing circumstances in order to achieve satisfaction of his wants. A goal which is wanted may be said to determine, as by final causality, the action which thereby occurs by "rationalizing" that action and therefore being that for the sake of which, i.e., for the attainment of which, the performance of the action is chosen or willed. Normally the term 'final cause' seems to suggest a goal achieved, the final state reached by (successful) rational action. However I have, earlier, stressed that wanted end-states only rationalize action conjointly with beliefs about present circumstances. The ambiguity of the term 'final cause', as well as the ambiguity of the term 'causality' when used by itself, suggests that it is wise to avoid this phrase in favour of a suitable substitute.

'Rational determination' seems to constitute a suitable substitute and will be used exclusively from here on. It is this notion which provides us with conceptual tools for understanding the human will in a way that is contra-causal yet deterministic.

The notion of contra-causality involved in this conceptual framework has particular importance in respect to the physical (causal) necessities which must apparently be overcome by an active originative mind if its activity is to manifest itself in overt bodily action. This notion appears, at first sight, to be quite problematical. As noted previously, the very shape into which Popper, for one, has cast the issue of free will is that of making sense of it in contrast to a thesis of physical determinism: the thesis that the world is a physically complete or physically closed system, a system operating ". . . in accordance with definite laws of interaction that do not leave any room for interaction with, or interference by, anything outside that closed . . . system of physical entities".[9] It has occasionally been argued that statements about human behaviour are not reducible to statements about mere bodily movements.[10] Such an argument might seem to provide a basis for establishing that the causal determinism of the entire physical realm would not of itself entail the causal necessity of human behaviour. However, as G. J. Warnock has pointed out[11] if we assume that mechanistic determinism

does hold exhaustively for the physical realm, then the physical description alone of all the particles of matter which are involved in some sequence of events which happens to constitute the performance of a human action, though it may not necessarily entail that some particular determinate human action have been performed, nevertheless does entail that no human action could have been performed in the situation which would have been incompatible with the physical account of how all the particles of matter behaved. If causal determinism is an accurate account of regularities among physical phenomena, particularly at the level of neurophysiological and muscular events, then it is, at first, hard to conceive how any activity at the level of mental phenomena could nullify what would of causal necessity have to be an exhaustive causal determination of the physical events which are constituent in, or amount to, human actions. If contra-causality is to pertain to human behaviour in any way at all, then it must be true that the physical entities which participate in the events of human behaviour (including not only the bodies of human beings but as well all the physical entities the physical state of which is interfered with as a result of human activity) be such that at the time of the occurrence of any particular human action, there is not necessarily only one state or event which can ensue, all things being considered.

The solution to this problem consists not in

arguing that human behaviour is not exclusively physical, but rather in pointing out certain features about the nature of physical theories to which attention was drawn in chapter two. There it was suggested that, for all practical purposes, theories must be regarded as being applicable to limited systems of entities or to systems of entities in certain limited respects. Theories according to which things in their specifically physical dimensions are subject to certain laws of causal necessity do not pertain to changes and regularities obtaining externally to the relevant system, in this case, the physical realm. As Stephen Körner has pointed out, the application of a theory to an empirical situation does presuppose--though does not guarantee--that the features of the empirical situation which have no counterpart in the theory are not only neglected, but are negligible--not only theoretically irrelevant but, in that situation, really ineffective.[12] In any case, it is the application of such theories, and not the theories themselves which presupposes such "closure" of the physical system. As Körner notes, physical theories by themselves do not logically imply that there are or that there are not what Körner calls "physically effective and physically independent choices".[13]

Suppose that all physical entities are thought of as comprising a system to which causal laws of motion and change are applicable and fully explanatory just so long

as there is no interference with the workings of the system by entities or forces not themselves part of the system, i.e., by entities or forces not themselves physical. Within the context of such a physical system, a regulatory principle such as that of Inertia in Classical Mechanics may be regarded as establishing, fundamentally, and in general, what is to count as a change of state within the system, especially among restricted relatively autonomous subportions of the entire physical realm. Such a principle implicitly establishes what is to count as interference with, or change of, the state of the physical system as a whole and the state of relatively autonomous physical subsystems by explicitly characterizing the states of affairs which would obtain in the absence of such interference.

The importance of this sort of principle to an understanding of the possibility of contra-causal activity lies in the fact that the principle seems to place no restrictions whatsoever on the nature of the possible physically disruptive entities or forces, with the minimal exception that they must be mechanically efficacious, which is to say that they must be capable of interfering with the movement and spatial location of physical entities which would have occurred (as defined by the Principle of Inertia) in their absence. The Principle of Inertia alone does not rule out the possibility that it may be mental phenomena which are thus mechanically efficacious. And if mental

phenomena need not be included in the system to which causal laws of motion and change are thought applicable, then there need be no systematic presumption that the occurrences of such disruptive mechanically efficacious mental phenomena themselves constitute changes of state which require further explanation in terms of causal regularities, i.e., in terms of their own causal antecedent conditions. The occurrence of those mental phenomena, even if they are mechanically efficacious, is a matter on which systems of physical laws are, as Körner claims, silent.

I have been suggesting that, in order to sustain a thesis of contra-causality which could be applicable to human behaviour, the mental phenomena or mental entities which are to be legitimately thought of as being mechanically efficacious must be conceived of as being excluded from the system of physical entities to which causal laws of motion and change are applicable. The exclusion must not, of course, be taken in a temporal sense, nor, for that matter, in a spatial sense, but rather in a logical sense. If physical entities could admit of changes of state which were caused mentally (and hence occurred contra-causally in all physical respects), then those changes of state themselves might have to be thought of as being both mental and physical. This is to say that, whatever form such changes could take, they would necessarily have to be such as to violate neither physical laws nor, if they obtain,

mental laws. But if this is to be possible, then it must be true that physical laws and mental laws are satisfiable simultaneously, that is, by the same events.

If rules of rationality indeed constitute a distinctively mental system of laws applicable to human behaviour, and if the physical realm is indeed causally deterministic, then in order that physical laws and mental laws be satisfiable simultaneously, it must be possible for the same events to be simultaneously parts of two formally different potentially recurrent unified sequence patterns. And in order for this to obtain, it would seem that two requirements must be met. Firstly, it must be in virtue of some determinable respect that an event belong to a specific potentially recurrent unified sequence pattern, so that it would be in virtue of different respects that the same event could simultaneously belong to formally different patterns. Secondly, neither the complete set of physical laws that obtain (whether they are known as such or not) nor the complete set of mental laws that obtain must be such as to restrict, in every respect, the possibility of what can occur in any given instance to necessarily only one determinate state or event, all things being considered. It would seem that this condition is indeed satisfiable by any set of laws which does not logically exclude the possibility of interference with the activities of the system to which that set of laws is applicable, by entities or

events which are external to that system. And no set of laws could logically exclude such a possibility unless they also included a law to the effect that there existed only the entities to which that set of laws applied and nothing else--a very strange law indeed.

A classical article in defense of soft determinism was entitled "Free Will as Involving Determinism and Inconceivable Without It".[14] This dictum might be used to represent my view, though not in terms of the thesis which this title was originally used to introduce. In the present context, such a phrase would be used to mean that there could be no choice of means appropriate to accomplishing certain ends if human actions did not themselves constitute reliable causally efficacious conditions in the world, effecting the changes that human beings want. As Gilbert Ryle has put it[15], "Predictability is a necessary condition of planning".[16]

A theory of agent-causality can thus be reconciled with a theory of causal determinism in regard to the physical realm simply by supposing that the contra-causal physical events are determined at least in part by conditions lying outside the physical realm. The notion of agent-causality is perfectly consistent with this characterization of contra-causal efficacy. The integrity of the laws of physical causation in this way remains inviolate. Furthermore they are supplemented with laws of

a different realm rendering the whole account a more nearly exhaustive explanatory framework for the entire world.

Such an account is clearly one of mental-physical interactionism, and one which is further complicated by the presumption of two different forms of efficacy for the two realms. I have no suggestions which will elucidate the nature of the mechanical efficacy which I wish to attribute to mental entities and events (nor, for that matter, of the psychological efficacy which I should equally wish to attribute to physical entities, as in perception). It seems deviously ad hoc to claim simply that such a relationship is sui generis and hence unanalyzable in terms of any other sort of notion. Rather than attempt to argue such a claim directly, I shall simply hope that my overall thesis regarding the need for such a framework is compelling enough in other respects as to make this claim seem less an evasion than a necessary last resort. One may cite Henri Bergson's point that such a view is no more absurd than the most radical of material theories: epiphenomenalism. For it is no more absurd to argue that consciousness can create movement out of a zero of kinetic or potential energy or by making use of this energy in its own way, than to argue that movement can create sensation out of a zero of consciousness.[17]

Certain additional compelling aspects of the theory presented here consist in its ability to survive

against several common attacks frequently leveled against libertarian positions. Libertarian theories which are presented as indeterministic accounts can easily be charged with a failure to take into account the large measure of predictive success which is, everyday, achieved with regard to human behaviour. C. A. Campbell, whose libertarian theory is of this sort, has attempted to meet this objection by arguing that only a very narrow range of human activity involves non-exhaustive determination by antecedent conditions, or, in Campbellian terms, involves the operations of a contra-causally free will.[18] Most particularly, these are actions occurring in practical situations in which conflicts arise in the agent's mind between what he conceives to be his "duty" and what he feels to be his "strongest desire", which is, in effect, the total expression of the conative and emotive dispositions comprising his character. In cases where the agent feels no such conflict, and acts in accord with what he feels to be his "strongest desire", then his behaviour, as well as the volition or decision which it manifests, is indeed exhaustively determined by the agent's character as so far formed along with the circumstances in which he has acted. Because these sorts of cases form the preponderant majority of cases, human behaviour clearly can seem to admit of a pervasive predictability. But it would seem that, according to Campbell, thorough examination should reveal

that predictability only occurs legitimately in those cases in which there is no conflict between the strongest desire and conceptions of duty.

I do not find Campbell's response (which I have taken the liberty of paraphrasing somewhat) to be completely adequate. It is not empirically obvious that situations involving duty-desire conflicts do not admit of legitimate predictability. Furthermore, by restricting the possibility of the manifestation of free will to just such cases, Campbell has unduly narrowed the scope of his views. If the contra-causal nature of mental events is indeed a possibility, then there is no reason to suppose that it actually occurs only in those cases where a sense of duty conflicts with the most strongly felt desire (although we might wish to say that its occurrence in such cases is more obvious, and more readily detected). Surely it is possible to imagine a similar case involving a conflict among mutually exclusive desires.

But more importantly, in contrast to Campbell, I think that actions performed for the sake of duty (as well as for other reasons) can be accounted for in a way which, without being causally deterministic, nevertheless still allows for the legitimate predictability of those actions. For I maintain that certain sorts of contra-causal phenomena, in particular, human decisions and the actions based upon them, may still occur as parts of potentially recurrent

unified sequence patterns in virtue of their causally undetermined respects. The important patterns in question are those which essentially involve practical syllogisms according to which behaviour is shown to be rational in the light of what is wanted and what circumstances are believed to obtain. A discussion of the more precise nature of the explanatory accounts, as well as a defense of the epistemological adequacy of the underlying generalizations shall be presented in the next chapter.

Another objection also raised against indeterministic libertarian theories is that the indeterminacy of any behaviour would render such events to that degree random and fortuitous, and not really acts of, or by, the person, mind, or self, to whom the actions would otherwise be attributed (on the basis, say, of bodily identity). Consequently any attribution of responsibility for such actions to that person, mind, or self would, to that extent, be incorrect.[19] But in light of the preceding discussion of this chapter and that of the preceding chapter, it seems possible to describe freely willed action as genuinely rational action and to regard genuinely rational action as action resulting from the present self-determination to act by an existent which is, thus, capable of self-determined rational activity. Therefore, genuinely rational behaviour can be explained as a manifestation of the self-determination, to act for the sake of reasons, by a being

or existent the essential nature of which must be such as to encompass, among other possible things: rationality, the ability to "take account of reasons", that is, the ability to reason correctly; and some degree of mechanical efficacy which is genuinely spontaneous as both rational and physically effective activity.

Whatever problems an "agent causality" theory might have, it at least seems capable of resolving the problem about the apparent non-responsibility of the person which seemed to arise on the basis of an assumption that contra-causal phenomena could only be random or fortuitous. For I have been suggesting that causally undetermined phenomena may yet be determined in another sense, namely, that of being rationally determined. If my arguments are correct, then there would appear to be no real problem of responsibility at all. Rational actions would, therefore, be such as to reflect the reasons for the sake of which they were performed and the essentially rational nature of the beings which (who?) performed them. Such a being, or agent, would be a non-physical enduring self, or mind, in a fundamental metaphysical sense.

A further objection to libertarian theories which is relevant to my deterministic version is raised by Broad.[20] A consideration of this objection will provide an opportunity for me to clarify one important feature of the theory which I am presenting. In some formulations of

libertarianism, as Broad characterizes it, the putting forth of a certain amount of effort in a certain direction at a certain time is determined in a unique and peculiar way, namely, by having among its determining conditions an agent or self considered as substance or continuant. Thus the series of events beginning with such a putting forth of effort begins with what Broad calls a causal progenitor, and which is a continuant and not an event. Such puttings forth of effort are neither completely determined, in the sense of containing an event as a cause-factor; nor are they undetermined, in the sense of occurring randomly and not "by" the self to whom responsibility is to be attributed.

Roderick Chisholm, for example, revives the medieval distinction between "transeunt" and "immanent" causation in order to clarify such a view in his own work: transeunt causation involves one event or state of affairs causing another event or state of affairs; immanent causation, on the other hand, involves an agent as causing an event or state of affairs.[21] Chisholm suggests that the question might be raised as to what immanent causation would consist of if there is no event other than some event, call it E, caused by the agent who could be described as making E happen. After all how would such a situation be distinct from the situation in which E just undeterminedly and randomly happens? This problem, by no means denied by Chisholm, is likened by him to a similar problem which may

be raised with regard to the notion of causation as it obtains among events. For with regard to transeunt causation we may just as readily ask "What is the difference between saying of two events, A and B, that A happened and then B happened--and saying that A's happening was the cause of B's happening? The difficulty is one of understanding the metaphysical nature of causation in either case; it is not a difficulty peculiar to the notion of immanent causation.[22]

The problem which Broad raises against the notion of immanent causation is that since the putting-forth of an effort of a certain intensity, in a certain direction, at a certain moment, for a certain duration is itself an event, or process, it is therefore subject to the same conditions which apply to every event. Most importantly, the total cause of an event or process which begins at a certain time must contain essentially ". . . another event or process which <u>enters into</u> the moment from which the determined event or process issues".[23] Broad is arguing that an event would not be determined to happen at a certain date unless the notion of date was applicable to its total cause. Yet the notion of date would apparently be entirely inapplicable to anything which is not an event or process, and hence to any part of the total cause of the putting-forth of effort which was not an event, in particular, the <u>agent</u> or <u>self</u>.

Broad suggests that the possibility of what he calls the "non-occurrent causation of events", as contrasted with ordinary (occurrent) causation, is mistakenly conceived as an account of behaviour which is determined by "Reason", "Principle", "Conscience", or "The Moral Law", notions which suggest a determination of behaviour by timeless propositions or universals or Platonic "Ideas". Such "entities" cannot literally determine behaviour, Broad argues, because they have no causal efficacy. He suggests that what determines behaviour in such cases would be a belief that a certain alternative was in accord with the moral law and a desire to do what is right, both constituting events and therefore proper cause-factors in the total cause which determined the agent's putting-forth of effort in accord with that moral law.[24]

The inadequacy of this view as a causal account of free, rational behaviour has already been discussed in earlier portions of this chapter. What is important in the present context is to indicate how an agent in the sense of a substance or continuant may properly be conceived as the cause of behavioural events which begin in time. It should firstly be noted, in accord with the discussion of "sets of causal conditions" in chapter two, that the agent as "cause" must be regarded as one among a set of conditions causing overt bodily behaviour. In this case, additional conditions which are necessary, but not causally sufficient for the

overt behaviour, are rather superficially characterized as the wants and beliefs of the agent. Bernard Berofsky has suggested, contra Broad, that since the action may be regarded as having such conditions as beliefs and wants which do provide temporal considerations, then it is no longer so much of a problem that "the cause", the agent, remains atemporal.[25] Without having to subscribe to Broad's conclusion, it may, I think, be said that this move will not do to avoid Broad's problem. For Broad is suggesting that there must be a sufficient condition for the occurrence of some event at a particular time. And this is to say that there must be a sufficient condition for the occurrence of an event at one particular time rather than at any other particular time. And for this degree of determination it is not sufficient that merely some of the conditions of the event are properly temporal.

The most viable solution to the problem consists in denying the timelessness of the agent or self. I suggested previously that in order to sustain a thesis of contra-causality which could be applicable to human behaviour, the mental phenomena or entities which are to be legitimately thought of as being mechanically efficacious must be conceived of as being excluded from the system of physical entities to which causal laws of motion and change are applicable. This exclusion, as I cautioned at the time, must not be taken in a temporal sense, nor, for that matter,

in a spatial, sense, but rather in a logical sense. Thus the exclusion of mental phenomena or entities from the applicability of causal (physical) laws entails nothing about the "where" or the "when" of mental phenomena or mental entities, and certainly does not entail that they are nowhere or nowhen.

To be properly said to "exist in time" it would seem that an entity or system must admit of change. The role of the agent in overt behavioural manifestations shall therefore have to be taken to involve change in the agent as one determining condition of overt bodily action. The change in question may be conceived, at its proximate end, as an act of will or volition which follows upon an agent's appreciation of the rationality of a particular course of action given what he wants to achieve and given the circumstances which he believes to obtain at the time. Such an "act", which is to be distinguished in kind from the acts of overt bodily activity performed intentionally and purposefully, may be knowable to us within experience as the exerting of effort in a certain direction, particularly in those cases where the doing of what we intend to do requires that major obstacles be overcome, psychological as well as physical. I have already suggested, in connection with a causal volition theory, that volitions may have the virtually exclusive status of theoretical events the occurrence of which is hypothesized as a general

account of the occurrence of human action. The need for such hypothesis arises, in the present context, out of just such problems as are raised by Broad. Thus a change in the self or agent, conceived as substantive or continuant, can be regarded as a temporal event that completes a set of conditions sufficient to bring about overt activity as an event or process having a definite beginning in time. The overt behaviour itself may be said to be "causally necessitated" in that its occurrence is not immediately the origin of new activity, but is rather the (merely transeunt) manifestation of such an origin. The act of will of the self may be said to constitute the origin of new activity, namely active response to the awareness of the rationality in terms of which the self conceives and intends its own action.

Broad's criticism may still be pressed against this account of acts of will. Probably this is what he intended since his concern is with the beginnings in time of those events which constitute the puttings-forth of effort, a characterization which can perhaps fit the experienced "feeling" of an act of will. But the solution to <u>this</u> problem is easy. The act of will is "determined" by the <u>prior</u> condition of the agent or self, namely, the appreciation of the rationality of a particular type of action, and this in turn is "determined" by the even prior conditions of belief and want. But as emphasized

throughout this dissertation, determination need not be causal determination, the mere transfer of a quantity of force or causal efficacy which are already given in the antecedent conditions. Determination, in itself, only amounts to regularity in accord with some general pattern, that is, nomological regularity. The logical contents of wants and beliefs determine an appreciation of the (logical) rationality of an action for an agent who is intrinsically rational and who is thus "determined" to act accordingly as a matter of rational necessity. Subsequently such an appreciation of the rationality of an action determines the occurrence of an act of will on the part of the agent to act in such-and-such a fashion. Again the determination is rational determination. I have argued that if one does not resort to theistic explanations, then the notions of rational determination and genuine rationality may only make sense if conceived as pertaining to the self-rationality, hence the original activity of an agent. No irregularity obtains. Nothing occurs without determining conditions. Determination simpliciter obtains throughout the process. Only the metaphysical interpretation of the nature of this determination differs from that which is commonly given with regard to a purely physical process.

I have been describing the agent, or self, as a substantival continuant. This notion will surely pose certain problems. C. A. Campbell describes such a being

as distinct from the states in which it manifests itself, and identical with itself throughout these manifestations.[26] Campbell himself broaches the problem of whether meaning can be given to such a notion. His response is essentially that self-consciousness is certainly a fact of experience, so that the terms of consciousness are capable of forming the framework in which the self can appropriately be described and conceived. We can therefore be aware of it as of such and such a nature. Conscious states of course differ through time, so description of the self in these terms will constitute an ongoing description of the self in terms of differing states or manifestations. But Campbell suggests that the sameness of the substantial self need not exclude all difference.[27] Now, this concept still seems problematic. It would seem more adequate to posit the notion of a being which evolves dynamically over time in proportion as it determines itself, as a materially and psychologically "embodied" self, to be a rational person in the world. The notion of the "self" may be said to constitute a being potentially capable of becoming rational, and ultimately self-determining to be rational as a public person. The more the self manifests its potentiality for original rational activity, the more it changes at least in becoming more defined and characterized as a public person.

At this stage of the formulation of agency-theory, the most important things which we can say about the concept

of an "agent" are not the intrinsic characteristics which we assign to an agent, but the role which such a conception plays in resolving such problems as that of accounting for moral and rational responsibility--problems which the only major form of alternative theory, namely, causal determinism, seems inadequate to resolve. The notion of an essentially creative self-rational being certainly seems capable of resolving those problems. It is easier to formulate this view as one attributing continuance to a hypothetically posited underlying substance regarded as source or origin of the initiative and rationality thus presumed to characterize behavioural sequences, rather than to take an alternative tack and suppose, as others have done[28], that we are dealing merely with a question of the applicability of two mutually non-reducible conceptual frameworks (the causal and the rational, or purposive, or teleological, or problem-solving) for understanding the data of human behaviour. Furthermore, the notion of a self, or agent, appeared to have distinct value in enabling us to understand how human activities could be characterized as meaningful given the assumption that the meaningful interpretation of those activities is not assigned from an external source. It is hoped that considerations such as these will make the metaphysical account of rationality and agency presented in this chapter seem generally compelling even despite possible flaws in the precise conception of

such a view as it is presently advanced.

One final point should be noted. The self-rational agency theory may even be seen to have a dimension of empirical "cash value", albeit indirect. Because the meaningfulness of human activity is self-originated, it must be discovered by the student of human behaviour rather than assigned, as is the case with regard to the interpretation of a mechanical system as constituting an adding machine. Thus in order to understand human behaviour in terms of its rationality, and not simply as a collection of merely deterministic regularities, it must generally be understood in terms of the meanings which it has for the behaving organisms. This means that human beings are the primary source of the particular concepts and meanings in terms of which their behaviour is to be conceived and explained. Of course, human beings qua students of any phenomena make the final decisions as to the particular concepts and meanings in terms of which to formulate their theories. But in regard to the special case of the study of human behaviour, the particular concepts and meanings in terms of which the theories are to be formulated must originate with the concepts and meanings in which human beings conceive their own behaviour qua behaving entities in the world, i.e., qua objects of study rather than qua students. Such personal meanings are the primary guide as to what the antecedent conditions, i.e., the particular

wants and beliefs, for a given action may be in any particular case. Such a methodological implication does not seem necessary on a merely causal account, for there is no causal reason to suppose that it would bear any fruit in terms of explanatory success. Consequently the practical success of this method of studying human behaviour can constitute, I submit, an indirect long-range over all empirical test of the metaphysical views presented in this dissertation, and one which, I should think, is already partly confirmed by current social scientific methodology,[29] not to mention the commonsense manner of understanding human behaviour.

FOOTNOTES

[1]Donald Davidson, "Actions, Reasons, and Causes", The Journal of Philosophy, vol. LX (1963), p. 51.

[2]Op. cit., p. 80.

[3]Cf. "Reasons and Causes", pp. 12-17.

[4]Ibid., p. 16.

[5]"The Case for Determinism", in Sidney Hook, ed., Determinism and Freedom in the Age of Modern Science (New York, 1958), pp. 10-12.

[6]Ibid., pp. 13-15.

[7]This theory draws upon ideas which are to be found in the following accounts: Thomas Reid, Essays on the Active Powers of the Human Mind (Cambridge, 1969), reproduced from The Works of Thomas Reid, vols. III & IV (Charlestown, Mass., 1815); G. F. Stout, Mind & Matter (Cambridge, 1931); C. A. Campbell, On Selfhood and Godhood (London, 1957); C. A. Campbell, "Is 'Freewill' A Pseudoproblem?", Mind, vol. LX (1951), pp. 441-465; Roderick M. Chisholm, "Freedom and Action", in Keith Lehrer, ed., Freedom and Determinism (New York, 1966), pp. 11-44; Roderick M. Chisholm, "He Could Have Done Otherwise" in Myles Brand, ed., The Nature of Human Action (Glenview, Ill., 1970), pp. 293-301; John W. Yolton, "Agent Causality", American Philosophical Quarterly, vol. 3 (1966), pp. 14-26; Richard Taylor, Action and Purpose (Englewood Cliffs, 1966); Richard Taylor, "Determinism and the Theory of Agency", in Hook, ed., op. cit., pp. 211-218. See also K. W. Rankin, Choice and Chance (Oxford, 1961).

Further sources which, as contra-causal accounts of human behaviour, have influenced me even though they lack

an accompanying metaphysics of agency are the following: Charles Taylor, The Explanation of Behaviour (London, 1964); Charles Taylor, "The Explanation of Purposive Behaviour", in Borger and Cioffi, eds., op. cit., pp. 49-79; William Dray, Laws and Explanation in History (Oxford, 1957); William Dray, "The Historical Explanation of Actions Reconsidered", in Sidney Hook, ed., Philosophy and History: A Symposium (New York, 1963), pp. 105-135; A. I. Melden, Free Action (London, 1961); Theodore Mischel, "Psychology and Explanations of Human Behavior", in Norman S. Care and Charles Landesman, Readings in the Theory of Action (Bloomington, 1968), pp. 214-237; D. W. Hamlyn, "Behavior", in V. C. Chappell, ed., The Philosophy of Mind (Englewood Cliffs, 1962), pp. 60-73.

[8]In "Postscript" to D. F. Pears, ed., Freedom and the Will (London, 1965), pp. 136-137.

[9]See footnote #3, chapter 5.

[10]Cf. Melden, op. cit., esp. chs. 3, 4, 5, and 8; G. J. Warnock, P. F. Strawson, and J. F. Thomson, "Determinism", in Pears, ed., op. cit., pp. 48-68; Richard Taylor, Action and Purpose, esp. chs. 4, 5, and 7.

[11]"Actions and Events", in Pears, ed., op. cit., pp. 76-77.

[12]Experience and Theory (London, 1966), p. 221.

[13]Ibid., p. 222.

[14]R. E. Hobart, Mind, vol. XLIII (1934), pp. 1-27.

[15]The Concept of Mind (New York: Barnes & Noble, 1949).

[16]Ibid., p. 81.

[17]Time and Free Will, trans. by F. L. Pogson (London, 1910), p. 152.

[18]"Is 'Freewill' a Pseudo-problem?", pp. 460-461.

[19]Cf. Hobart, op. cit., pp. 76-77.

[20]Op. cit., pp. 156-158.

[21]"Freedom and Action", p. 17.

[22]Ibid., pp. 20-22.

[23]Op. cit., p. 157.

[24]Ibid., p. 158.

[25]Determinism (Princeton, 1971), pp. 64-65. Berofsky, however, rejects agent-causality on other grounds, particularly its alleged pointlessness. But his reasoning is peculiar. He appears to be ready to allow for the possibility of "free action" for which there is simply an absence of sufficient condition and an absence of all features--e.g., passivity, surprise, unintentionality--that are mistakenly supposed to be present when a sufficient condition is denied (p. 70). This seems to amount to plain indeterminism and is hence subject to all the difficulties of that view as an account of responsible, rational human behaviour.

[26]Cf. On Selfhood and Godhood, p. 74.

[27]Ibid., pp. 82-83.

[28]Cf. Charles Taylor, "Mind-Body Identity, A Side Issue?", The Philosophical Review, vol. LXXVI, no. 2 (April 1967), pp. 201-213.

[29]For such an interpretation of social study, cf. Peter Winch, The Idea of a Social Science (London, 1958).

VII

THE SCHEMA OF RATIONAL EXPLANATION, QUALIFIED

In chapter one, I set out the following schema for the explanation of human behaviour:

People do what is appropriate for them to do (i.e. people act rationally).	X wants E. X is in circumstance C. If an agent wants E, then if an agent is in circumstance C, an appropriate thing for that agent to do is A.

∴ X does A.

This schema is expressed in terms of a rationalistic, and, I shall argue, non-causal interpretation of the pertinent behavioural sequence. By a "rationalistic interpretation", I mean that the sequence is conceived as one in which the important source of change is an essentially active, creative agent who is acting for the sake of (that is, so as to achieve) wanted ends. In this chapter, I shall offer some qualifications in the light of which the above schema and the entire "framework of rationality" shall be understood.

As argued in chapter two, alternative metaphysical interpretations may be given for the purely deterministic regularities which obtain among any realm of phenomena. The deterministic regularities themselves do not uniquely select a single metaphysical account as the appropriate one. Therefore the above schema of the explanation of human behaviour, because of its formulation in terms of what I shall call a "metaphysics of rationality", a rational (as opposed to a causal) metaphysics, is already biased in favour of an account for which arguments still remain to be presented. In order to avoid begging any questions, then, it is important to re-express the pertinent behavioural regularities in terms which are as neutral as is practicable, given the subject matter. The following qualifications on how the above schema is to be understood even from the point of view of a rational metaphysics will prepare the way for such a reformulation.

A distinction must be drawn between, on the one hand, judgments of the rationality or appropriateness of a certain type of action simpliciter, given a specified type of end-state and a specification of types of circumstances, and on the other hand, judgments of the rationality of a particular action for a particular individual who wants such an end-state and who is actually faced with such circumstances. The distinction is intended to point to a rather obvious fact that individuals, not uncommonly,

misperceive the circumstances confronting them or misjudge the nature of the means which would be necessary within a situation to achieve their wanted ends. Yet given such misperceptions and such misjudgments, such individuals may nevertheless act in ways which we should wish to regard as appropriate or rational. We might call this a rationality within the context of mistaken beliefs.

The above schema of the form of explanation of human behaviour in terms of a rational metaphysics, or what I shall call "rational explanation", has been formulated in such a way that the logically antecedent conditions involve a specification simply of the (relevant) circumstances which obtained at the time of action. However in light of the possibility of erroneous beliefs on the part of the agent regarding those circumstances, it becomes clear that the actual circumstances of action (in relation to which a certain sort of action is appropriate, given a certain want) may have no explanatory relevance to an action which is mistakenly performed under such conditions. The proper form of rational explanation must involve due recognition of the agent's own estimates of the circumstances and of the cause-and-effect relationships pertaining to those estimated circumstances, which are relevant to achieving what the agent wants. A further consideration will show that this qualification upon the nature of rational explanation is relevant not only to the explanation of

mistaken or misjudged action, but to the explanation of correct action as well. An agent who acts correctly relative to the (relevant) circumstances which obtain at the time of action and the cause-and-effect relationships pertinent thereto, is not acting rationally unless he is acting on his own appreciation of those circumstances and their action-relevance for him.

When it is judged that a certain type of action is rational in the context of certain sorts of circumstances, given a specified type of wanted end-state, the rationality thereby attributed to the action is purely a matter of the logical relationship, the relationship of appropriateness, which that type of action bears to the other types of factors involved. The appropriateness of action is a matter of the logical relationship between a possible action, a goal or possible state of affairs, and a set of envisaged circumstances--the intentional properties of belief events, states or dispositions. The model for the understanding of human behaviour in terms of its rationality must be one which accords the logical relationship between the intentional properties, or "contents", or "objects", of beliefs and wants an essential explanatory relevance. This logical relationship, of making appropriate a certain possible course of action, makes that possible course of action appropriate for any agent who would happen to have just those beliefs and just that want.[1]

The event of explaining something is an event which must involve the invoking of all and only such factors as could have in some way contributed to the occurrence of that which is to be explained. This applies equally with regard to the event of explaining the action of a particular person. That an action is appropriate given a set of circumstances and given some specified end-state is not explanatorily relevant with regard to the occurrence of the action unless the appropriateness of this action in some way participated in bringing about its occurrence. An ordinary physical process which culminates in some specific end-state may be said to have begun with an event which, given the circumstances at the time of the onset of the process, was <u>appropriate</u> for bringing about the specified end-state. But physical processes are not said to be rational in this practical sense.[2] Something more is required. That "something more" may be approximately characterized as something involved in the process of change which is capable of envisaging the circumstances and the possible end-states which can ensue, of taking account of them, and of participating in changes which reflect this account taken of those circumstances. A ready way in which we can conceive of an entity "taking account" of circumstances is for that entity to have <u>beliefs</u> about those circumstances, and to want some specific end-state. Let us call an entity which can thus have beliefs and wants an

"agent".

Consequently the logically antecedent conditions in the rational form of explanation must pertain to beliefs and goals which the agent in question actually had at the time of action in order that the argument constitute an adequate explanation of that agent's behaviour (in terms of its rationality). And this point holds whether or not the agent was mistaken in his beliefs about existing circumstances and the causal consequences of various possible actions amid those circumstances. The behaviour of even an agent who acts rationally in given circumstances is still only entitled to be explained in terms of its rationality if, at the very least, the agent in question correctly believed that circumstances did obtain which would enable him to achieve what he wanted and correctly judged the causal activities, that is, the actions, which he would have to perform to achieve it. So the more accurate form of rational explanation, whether in the case of mistaken (i.e. inappropriate) actions given actual circumstances, or in the case of correct (i.e. appropriate) actions given actual circumstances must take due account of how the agent himself envisages his situation in order that his behaviour in particular can be satisfactorily explained.

These considerations suggest that a proper rational explanation cannot be given for behaviour unless evidence is found regarding the nature of an agent's actual

beliefs concerning the circumstances of his situation. Clearly there are many instances in which nothing like direct independent evidence--such as verbal or written commentary by the agent himself--is to be had regarding such matters. It thus seems that there are important limitations upon the practical applicability of the form of rational explanation which depends on an appeal to such conditions. This problem might, for all practical purposes, be circumvented by reliance upon a certain heuristic assumption about human beliefs which can direct attention to sources of indirect evidence concerning the nature of the pertinent beliefs of the agent. This is the assumption that, unless positive evidence proves otherwise, or unless behaviour still remains unintelligible on these grounds, human agents whose behaviour we seek to explain, are to be regarded as perceiving their circumstances as accurately as we, the explainers, are able to perceive them or otherwise determine them to be, and also as believing the action-relevant aspects of causal order to be as nearly the truth as we, the explainers, are capable of determining them to be. This assumption functions only as a counterweight to the practical problems of research into human behaviour. It is more likely, in researching the "Why?" of human behaviour, that direct evidence will be had as to the nature of the actual circumstances surrounding behaviour than will be had concerning the agent's apprehension of

those circumstances. The most that we are able to hope, on the basis of such evidence, is that the agent was in some "normal" state of mental outlook and awareness, and that he therefore regarded his situation as accurately as we, on the basis of our evidence, are able to regard it. It must be stressed, then, that the form of rational explanation, when expressed in terms of the actual circumstances of action, rather than in terms of the circumstances as believed by the agent to have obtained, is a bastard product of the ideal form of rational explanation coupled with a pragmatic limitation on our abilities to glean ideally relevant evidence.

Suppose that we are unable to grasp some action as being appropriate in the circumstances as we determine them to be, and subsequently suppose the agent to have envisaged them. There is nothing for it, in such cases, but to seek further evidence which would suggest the plausibility of attributing different beliefs--or even different goals--to the agent in question, in terms of which the action in question (which we presumably know to have been performed; this is our anchor in the sea of data) can then be appreciated as being appropriate. In some cases this may lead to our postulating (what we regard as) certain outlandish or totally erroneous beliefs on to the agent. Such hypotheses might be confirmed if, for example, a whole repertoire of the agent's behaviour can be seen as

appropriate only by assuming such beliefs to obtain on his part. Yet we may wonder how the outlandish or erroneous beliefs could have been believed by the agent in the first place. And as the explanation of an action in terms of such beliefs may not seem satisfactory or complete until some further explanation is forthcoming which will account for the existence of such beliefs on the part of the agent. We might thus sometimes not feel that we had grasped an agent's behaviour as being rational until this additional understanding is obtained.

However it is not clear to what extent the having of beliefs on the part of an agent is itself to be explained in terms of its rationality. Certainly parallels obtain between the drawing of conclusions and the performing of an action: both can be regarded as a "doing" for, or against, which reasons can be given.[3] Thus each is guided, in general, by a certain kind of objective: for belief, the objective is knowledge of the truth; for action it is the attainment of some sort of wanted end. As practical reasoning is generally involved in the determination of what is to be done, and the practical syllogism can display the logical form of such reasoning, so, in the special case of belief, theoretical reasoning is involved in the determination of what is to be believed, and the theoretical syllogism can display the logical form of this sort of reasoning. Such parallels help to unify under a single

formal account a variety of different things about human beings that we seek to understand, and thereby strengthens the acceptability of this account for either sort of case. Beyond these remarks, the parallel shall not be pursued any further in this context. In any case, the important point is that whether or not we also demand an explanation for beliefs in order to feel fully satisfied that those are the beliefs which correctly explain a given action seems to be a practical context-bound affair which will depend on such theoretically irrelevant factors as the strangeness or erroneousness of those beliefs. This is not a philosophical problem affecting the adequacy of the form of the rational explanation of <u>behaviour</u> as herein proposed.

Further clarification should be given regarding the sorts of antecedent conditions which are invoked by the rational explanation of behaviour. At any one time, an agent faces an environmental totality which is indeterminately vast as regards the possible actions which it allows for a given agent. In accord with the way in which I am attempting to explicate behaviour, it would follow that only such possible actions need merit our attention as would enable the agent to achieve wanted ends.[4] Hence we need not concern ourselves with actions which are irrelevant with regard to anything which the agent wants to achieve. But we may still have a very sizeable class of actions remaining, each of which an agent could in principle have

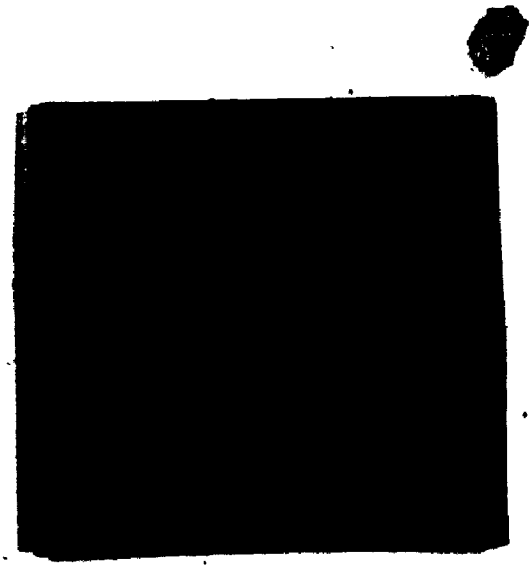
grasped to be possible for him to perform at some given time in his life, in order to achieve something that he wants. Which action, or reasonably small set of actions, is appropriate for an agent to perform can only be determined by considering a number of important factors. Firstly, human beings can usually be relied on to be rather complex beings characterized by a totality of wants--goals, desires, principles, etc.--that may, very superficially be regarded as comprising a hierarchy. The principles of such hierarchical ordering encompass such factors as the degree of importance attached to the want, the degree of passion attached to it, and the extent to which some wants are themselves subservient to others in being wanted intermediately only because they are the means necessary to achieving those other ends. This fixed concatenation of wants may be said to form part of the ongoing identity of a human agent, centrally determining all aspects of the agent's behaviour and hence the character of the public person which that agent becomes through living his life. This hierarchy has stability, but is also somewhat dynamic, changing over time. Furthermore not all wants are equally pertinent to any given situation. A situation may only be relevant to (by allowing for the possibility of) the attainment of a small sub-set of the agent's wants. Thus, in specific situations, the "standing" want-hierarchy of the agent will be translated into a situationally-relevant

ordering on the basis of such factors as the degree of probability of actually attaining each want through potentially effective action at that time, and the risk of harm which may occur as an additional consequence of performing such actions. In sum, not all wants are valued equally and not all wants are equally satisfiable in any given situation.

The reconciliation of all these factors for a given agent on a given occasion is a matter which can safely be regarded as beyond the scope of this dissertation. Recent literature in what is called "decision theory" has been devoted extensively to solving the various sorts of problems alluded to above in the hope of providing a method for arriving at unique solutions to the question of how finally to act in a complex situation.[5] It may be noted that all the above problems are further compounded by the fact that an agent may "define" his situation differently.[6] That is, an agent may attempt to match (appropriate) action to a situation regarded as just obtaining at that moment and ending in the very near future, or, rather, to a situation regarded as having begun more remotely in the past (that is, as encompassing "circumstances" which obtained in the past) and perhaps not ending until some more distant time in the future. At any one time, the "present situation" may thus be quite variously conceived, and this conception will importantly determine the way in

4 5

OF/DE



which the agent's standing hierarchy of wants will be translated into a hierarchy of appropriate actions for a given time. It may also be that the agent's decision about which situation he regards as action-relevant at a time may not be made independently of the action he decides to perform and the objective he decides to pursue. One and the same decision may constitute a commitment regarding both these matters. Nevertheless the two elements of the decision are conceptually distinguishable, and therefore distinguishable for the purposes of understanding and explaining the behaviour which has been performed. One element is cited as the agent's (situationally-relevant) want, and the other is cited as the agent's beliefs regarding the situation in which he acts, however reasonably or arbitrarily the agent arrives at the decision which he makes in this regard.

These plentiful factors which comprise the agent and his "situation of action" at any given time provide more difficulties for the explainer of the agent's action than they appear to do for the agent-in-action. Except for those infrequent cases requiring extensive deliberation, it does not normally seem that we are faced with so vast a plethora of circumstances to reconstruct and reconcile in deciding what (is the appropriate thing) to do. Perhaps this is the result of previous decisions to act in ways which carried long-range commitments to patterns of

behaviour that we continue to execute without redeliberating afresh the nature or appropriateness of what we do. Perhaps through training or laziness we relegate many of our activities to the status of habits so that our responses to rather standard well-defined sorts of situations follow as a matter of thoughtless course upon the occurrence of those situations.

As a consequence of such tendencies, the job of the student of human behaviour is made all the more difficult, for he cannot even claim to be re-constructing, as an explanation of some action, a picture of the agent's outlook or point of view pertaining to the performance of that action. For beliefs and wants may thus be relevant to an agent's situation in a way that is not always a matter of conscious awareness for an agent. Even in principle, it seems, the agent at the time of action need not be capable of giving us all the information which we should need regarding his beliefs and wants in order to reconstruct an account of his behaviour in terms of its rationality. How then can it be supposed that a rational explanation can always be given in principle for even such behaviour as does <u>not</u> follow upon a conscious awareness on the part of an agent of all the factors requisite to constituting, in him, a rational (in a practical sense) frame of mind as the basis for action?

Some very interesting suggestions are made by

Karl Popper[7] which provide the basis for a solution to this problem. Popper suggests that human behaviour, as well as the behaviour of all organisms, is to be explained and understood as the activity of problem-solving, where the "problems" involved are problems in an objective sense: that is, they are to be hypothetically reconstructed by hindsight.[8] Since this theory is to be applicable to the behaviour of all organisms, its adequacy to the subject matter clearly requires that conscious awareness of a problem as such need not be a necessary condition for the operation of problem-solving activity. Each organism, according to this theory, is to be regarded as always engaged in problem-solving, reacting to new and old problems by more or less chance-like trials which are eliminated if unsuccessful.[9] The "problems", in the objective non-psychological sense which the organisms are to have been constantly engaged in solving are to be reconstructed ex post facto, apparently only on the data pertaining to what has in fact occurred. Popper notes that even scientists, perhaps the most consciously articulate problem-solving beings, are not always fully aware until afterwards of the problems they solve by their deliberative activity. Hence he sees no objection to supposing that an amoeba is also capable of problem-solving activity, though it will not be assumed that the amoeba was, or ever could become, aware of its problem. We are to regard the difference between the

amoeba and Einstein to be one of degree rather than kind: unlike the amoeba, Einstein has a conscious critical attitude toward his own ideas, and by means of them, toward the problems he faces and the errors embedded in his tentative theories and hypotheses--a method which Popper calls a faster, more efficient, exosomatic method of problem-solving hypothesis elimination.[10]

Popper is thus suggesting a form of understanding for a certain range of phenomena. If successful, this form of understanding will enable us to apprehend an important degree of regularity among the selected phenomena in terms of a single comprehensive structural account, and will lead us to discover further regularities among the phenomena that we might otherwise have missed had we not thought in such terms. It is not of particular concern, from the point of view of adopting an explanatory framework, that the organisms themselves do not have conscious awareness of the factors that constitute the explanantia of their activities for us. Physical entities similarly lack conscious awareness of the factors that constitute the explanantia for us of the changes which they undergo, yet we hardly regard physical explanations as defective on that account. The important consideration is the explanatory and predictive success and the theoretical satisfaction which a given structure of explanation affords us in regard to a range of phenomena. Hence Popper's thesis can be thought of as

based on the argument that an account of behaviour in terms of problem-solving, and attendant notions, will afford us greater theoretical satisfaction and pragmatic success than any other sort of account, or, at least, than any other structure of explanation presently regarded as relevant to the subject matter.

However there seems to be a problem for Popper's account, which is hinted at by the statement that physical entities similarly lack conscious awareness of the factors that constitute the explanantia for us of the changes which they undergo. Why suppose that the same structure of explanation as is to be applied to the behaviour of living organisms should not also be applied to the purely material non-living universe? Why could not some purely physical entities be regarded as constantly engaged in more or less chance-like trials which are eliminated if unsuccessful, and thus as progressively working and evolving toward the solution of various problems? Popper does point out that there are purely physical analogues to the random chance-like trials which, among behaving organisms are "plastically" controlled or directed "from above" from controlling systems which exert such control "softly" over the randomly active subsystems that engage in the trials.[11] However he does not indicate that the purely physical cases are also to be thought of in terms of the notions of problem-solving, tentative hypotheses or theories, and error elimination.

They are merely adduced by way of showing the plausibility of a notion requisite to the understanding of problem-solving activity, namely, that of a hierarchical system of plastic controls: controlled sub-systems which are limited by, and yet have their own effect upon, their controlling systems. But if purely physical analogues can be found for that which, in the case of living organisms, constitutes problem-solving activity, then there seems to be no reason not to regard the same concepts to be applicable to those purely physical processes.

A plausible way of ruling out the applicability of notions of problem-solving to purely physical systems--something which I should think that we intuitively want to do--would be to postulate, as a necessary condition of such activity, that the entities involved somehow "take account" of, or envisage, or intend, their problem to be their problem. The same sort of condition was postulated earlier (chapter four, p. 140) as a differentia for goal-directed systems. In fact the two sorts of accounts of the understanding of behaviour may be seen to dovetail in the following way with regard to human behaviour: the problem in terms of which human behaviour is to be understood as the solution is the practical problem of achieving wanted ends. And the "taking account" of, or envisaging, or intending of such ends, and hence such practical problems is established with regard to human behaviour, in a general way, precisely

through the evidence which we occasionally collect regarding the conscious awareness of themselves and their own conditions that human beings possess. With regard to types of entities for which we lack such evidence--that is for purely physical entities--there, we lack any independent reason, other than the process of change itself, to suppose that the end-state of that process, or the occurrence of the process itself, is intended or envisaged at the outset. The mere process of change alone does not appear to constitute a necessary sign of intention in any quarter.

However Popper would respond to this problem with regard to apparently totally non-conscious organisms, his account of the explanation of behaviour may still be considered relevant to an understanding of the behaviour of those paradigms of consciously aware organisms: human beings. We can now supply an answer to the previous question of how it can be supposed that a rational explanation can be given, in principle, for even such human behaviour as does not follow upon a conscious awareness on the part of the agent of all the factors on his own part requisite to constituting in him a practical rational frame of mind as the basis for action. The answer is that the postulation of such a complete frame of mind even where, in a particular case no independent evidence sustains such a postulation, is justified on the grounds that, in general, independent evidence (based on other instances) confirms

the view of human beings as intentionally goal-seeking, problem-solving agents. The extrapolation of such a view to those cases where such evidence is not in fact available is just a way of systematizing and unifying as much as possible under one explanatory rubric, the overall account which is to be given of human behaviour. The rational form of explanation is therefore justified in regard to cases in which the agent lacks conscious awareness of deliberative rational choice just so long as it is pragmatically successful in regard to those cases--i.e., enables explanation and prediction in regard to them--and so long as it is theoretically satisfactory, i.e., fits into a more inclusive unified theory which is adequate to the subject matter as a whole.[12]

One further clarification of the form of rational explanation is necessary at this point. The explanation is schematized as depending upon a law of the following form: "If an agent wants E, then if an agent is in circumstance C, an appropriate thing for that agent to do is A"; and the entire explanation is formulated in accord with a framework principle for the understanding of human behaviour which has been given the following expression: "People do what is appropriate for them to do". Neither the law nor the framework principle are formulated in terms of what is "the appropriate thing" for a person to do in a given situation. But if the rational explanation of action is

to exhaustively explain human behaviour, then some account must be taken, at some stage of the explanatory process, of what is the appropriate thing for a person to do. If not, then behaviour cannot have been exhaustively explained in terms of its rationality.

The following should make this difficulty clear. It has been noted that human beings are generally rather complex beings characterized by a totality of wants that may be regarded as comprising an ordered hierarchy. A given situation may be relevant to the attainment of more than a single of those objectives. I have noted the complexity of the "decision problem" which faces an agent in determining how to act in such a situation, and have suggested that a discussion of the method for reaching such a decision is beyond the scope of this dissertation. However this "decision problem" does point up a related point that is of revelance in the present context. This is the point that although there may well be a unique and decidable "most appropriate" thing for a given agent to do in a given situation, all things considered, nevertheless several actions (including the unique "most appropriate" thing to do) may each, relative to some different want or some different feature of the situation, be an appropriate thing to do. A rational explanation schema which does not incorporate a reference to the factors on the basis of which one among several appropriate things to do is finally selected

as the (appropriate) thing to do does not fully explain the occurrence of a unique action. The action is thus only partially explained, namely, in respect of its being one among a set of appropriate things to do one of which, never mind which exactly, was actually done. The action is precisely not explained in respect of its being the one which, in contrast to other appropriate things to do, was actually done.

Note that this is not the problem of how to explain, in terms of its rationality, why some unique action among several equally most appropriate things to do, is finally done. Suppose that it is most appropriate for an agent, at a given time, to save the life of someone who has just fallen overboard, and the most immediate thing to do is for the agent to throw the man a life preserver. Suppose also that there are three life preservers all equally ready to hand, and all equally sturdy and reliable. Shall our agent stand paralyzed with indecision because no rational basis exists for deciding which, among them, to throw to the drowning man? This problem shall have to be dealt with eventually, but for the time being I only wish to note that it is distinct from the problem which is presently in mind. The problem presently in mind can be illustrated by means of the following example: Suppose that at some given time, it is possible for an agent to achieve each of several significantly different things, though not any two together, and

each by means of some distinct performance on his part which excludes, for the time, the performance of any of those other actions. Relative to some different want on the agent's part, each of these actions is an appropriate thing to do. Consequently, for a given agent at a given time, several distinct laws regarding what is an appropriate thing to do (I shall, following William Dray,[13] be calling these laws "principles of action") are each pertinent. It is not clear yet why only one such law should be regarded as explanatorily relevant to the unique action that is actually performed.

The problem is resolvable in terms of the following considerations. I have already pointed out (in chapter one, p.13) that the covering law model shall constitute the working epistemological abstraction into which I shall endeavour to fit my account of human behaviour in terms of rationality. It was noted (chapter one, p. 16) that the covering law model is most clearly applicable with regard to explanations in which only one condition or one type of condition, or some very small set of different types of conditions, are cited as explanatorily relevant. But with regard to more complex cases, in which a wide variety of different types of factors all appear to have explanatory relevance, then the covering law model must be regarded simply as an idealized philosophical abstraction for representing the structure of explanation in such cases,

and one which does not represent all aspects of the dynamic explanatory process. In particular, since it seems that most generalities which we ever have occasion to formulate are economical in form, linking a rather small number of types of antecedent conditions to a type of consequent, then what the covering law model expressly fails to exhibit is the way in which we reconcile, in terms of weight and priority, a plurality of covering laws which are invoked with regard to such cases in which a plurality of types of explanatorily relevant factors can be discerned. As stated earlier (chapter one, p. 16), this process of reconciliation, for which no further generalities may be available in all cases as rules to guide us, requires the art of judgment.

Principles of action, each of which formulates an appropriate thing to do relative to a situation-definition and to a certain specified want, can be regarded as economical generalities that have a useful explanatory (and practical!) function precisely because of their real practical generality: they each pertain to a wide variety of actual instances. And in such instances in which only a single situation-definition and a single want on the part of the agent are relevant to an explanation of what the agent does in that instance, then for those instances no complex process of judgment is involved, and the covering law model of rational explanation most clearly depicts the

structure of the explanatory argument. However, analogously to complex physical explanations, it is with regard to those instances in which there are a variety of appropriate things for an agent to do, depending upon which want and/or which circumstance is cited, and hence a variety of pertinent covering principles of action, that the abstract covering law model is less adequate in schematizing the structure of the entire process of explanatory reasoning involved. Precisely what the covering law model fails to exhibit, though no more so in the case of human behaviour than in the case of complex physical explanation, is the process of judgment by which the various explanatorily relevant factors--and hence the various covering laws--are reconciled so as to determine why some unique event in particular has occurred. Thus what the covering law model of rational explanation fails to exhibit, because it is an idealized philosophical abstraction, is the process of reasoning by which it is determined what in particular is the appropriate thing for a given agent to do at a given time, this determination constituting a crucial judgment if an action is to be as fully explained as possible in terms of its rationality.

I have already indicated that a discussion of the process of reconciling all circumstances, wants, risks, and so forth in a given case so as to yield a unique determinable "thing to do" for any given agent on any given

occasion is beyond the scope of this dissertation, and shall be left to the "decision theorists". For present purposes, it shall simply be assumed that, for most cases, unique types of actions "to be done" are in principle, if not in fact, decidable. The focus of attention, for the purposes of this dissertation, shall be on the covering law explanatory schema in its idealized abstract form, that is, as most clearly relevant to the explanation of behaviour in the simple sorts of cases where only a small set of factors are regarded as explanatorily relevant, and, hence, only a minimum number of pertinent covering laws are invoked.

It should be noted that even in such simple cases, whether the explanation be of human behaviour or of any other sort of phenomenon, there is still a measure of judgment required in appreciating a given (simplified) explanation as adequate to the event to be explained: in particular, the judgment is required that no further factors are explanatorily relevant to the explanandum than the small set to which attention has been directed. And in the explanation of human behaviour, this judgment amounts to saying that what is an appropriate thing to do relative to a certain situation-definition and a certain want is, in this case, the appropriate thing to do all things considered. This final judgment closes the last vestiges of any "logical gap" that remained between the explanans and the explanandum of my covering law model of rational explanation.

The question arises then as to whether or not it would not be more satisfactory to schematize the principle of action involved in such explanation in terms of "the appropriate" thing to do. This formulation might be deceptive for the following reason. It may seem to suggest that the sorts of economical principle of action generalities which we carry around us as a standing repertoire for use in understanding human behaviour (and for our own practical decision-making!) are each of the order of "final judgments", not just true of situations ceteris paribus, but true of them absolutely. I do not intend to suggest that this is the case. Rather I wish to emphasize that what is an appropriate thing to do relative to certain conditions which may be discerned in a given situation may not be the appropriate thing to do at all in that total situation, depending on what other conditions are also present. Abstract principles of action, in other words, may each be regarded as obtaining ceteris paribus.

However if the distinction is kept in mind between the final implicit generalization which covers an explanation (of whatever sort) and the standing stock of generalities which help us to get about in the world, then there may be no confusion in schematizing the covering law model of rational explanation in terms of a law of the form: "If an agent wants E, then if an agent is in circumstance C, the appropriate thing for that agent to do is A". It must

simply be remembered that this generalization is not a standing principle of action but, rather, the implication of applying a standing principle of action to a unique situation coupled with the judgment that only that principle of action is relevant to understanding (in terms of its rationality) the behaviour that has occurred in that unique situation. Paralleling this change in the schema of the rational form of explanation, a corresponding change must be acknowledged in regard to the framework principle which itself defines the form that the explanation of behaviour in terms of its rationality is to take. This principle therefore becomes: "People do what is the appropriate thing for them to do".

One final comment related to this clarification remains to be discussed. Previously I distinguished the above problem of whether rational explanation had to be formulated in terms of "the" or "an" appropriate thing to do from the problem of how unique behaviour was to be explained if in fact it was one of several equally most appropriate things to do in a given situation. This problem, I think, can only be handled in the following manner. The explanatory model presented here is one pertaining to the understanding of human behaviour in terms of its rationality. Thus, in so far as there is a rational basis for the choice of one action rather than some other in a given situation, then to that extent the actual choice of that one action

(rather than the other) is comprehensible as such. But in so far as there is no rational basis for the choice of some one action rather than some other in a given situation, then to that extent the actual choice of that one action is not comprehensible in terms of its rationality. For just to that extent, the choice is not rational (in any practical sense).

This sort of explanatory situation seems capable of occurring in regard to two sorts of cases: one, the case in which a plurality of actions are equally appropriate ways of doing the same most appropriate thing to do, such as in the previously noted case of the agent who is to save someone's life and has a choice of three equally appropriate life-savers to throw overboard; and two, the case in which the sort of "decision problem" alluded to earlier, in which the action-relevance of various different wants and circumstances must be reconciled so as to determine a unique "thing to do", is undecidable. This might occur, for example, if there were no legitimate rule for quantitatively evaluating various different wants or circumstances against each other so as to establish a hierarchical ordering of appropriateness with respect to a unique case. If this is indeed a problem for decision theory, then the type of case in which this sort of problem would occur is one in which, for all intents and purposes, there may be said to be a plurality of equally relevant "most appropriate" things to do.

We do know that in such cases, at least in the first sort, decisions are frequently made. The agent (it is hoped) will throw one life preserver overboard. It is simply the case that we cannot fully explain in terms of its rationality the unique action performed in such cases, because the unique action performed is not a matter of rational choice in all respects. This is a limitation not upon the covering law model of rational explanation, nor upon its supporting theory, but rather upon the nature of the subject matter itself: the extent to which human behaviour can be rational is limited by the extent of human wants. Beyond these limits, it simply does not matter, from our point of view, what we do.

Suppose, for example, that we are discussing the behaviour of the agent who is rescuing the man overboard. You may ask, "Why did he throw the #2 life preserver overboard?" My answer is likely to be, "In order to save the life of R who had just fallen overboard". Suppose that you persist, "But why did he throw the #2 life preverver overboard?" I would likely interpret such a question to mean, "Why did he throw a (any) life preserver overboard rather than commence with some other technique for rescue on water?" and shall respond to the effect that when someone falls overboard, the first thing to do is to assure that they have some means of staying comfortably afloat in the water while other rescue procedures are put into operation.

Suppose that you are still persistent: "But why the #2 life preserver rather than the #1 or the #3?" Assuming that all the life preservers are equally ready to hand and equally suitable for the purpose, the question is frivolous. It hardly seems to matter which life preserver is thrown so long as one of them is thrown. Consequently it hardly seems reasonable to ask for an explanation in terms of its rationality of that which happens to make no rational difference.

Such aspects of behaviour, which are inexplicable when understood in terms of the rational framework, may be compared to aspects of purely physical events which are similarly inexplicable when understood in terms of some (other) pertinent form of deterministic theory. Thus in chapter two it was suggested (page 60) that the application of theories of deterministic form in our grasp of the world involves at least two non-determinate dimensions. Firstly external interference with the workings of an abstractly isolated system of theoretically-relevant entities cannot always be ruled out. Secondly (page 62), because measured values are never mathematically precise numbers, but rather mathematical intervals, they will diverge from the precise results calculated by the use of the theory. I suggested that the general applicability of some deterministic theory to a range of phenomena need not be impugned by the existence of such undetermined

experiential "fringes". Similarly, I would argue, the applicability of the (deterministic) theory of rationality to the phenomena of human behaviour need not be impugned by the existence of undetermined "fringe" aspects of those behavioural events. We do not need to understand absolutely <u>every</u> aspect of some event in order to have <u>any</u> understanding and knowledge of it at all.

In the light of the foregoing discussion, the schema for the explanation of human behaviour in terms of its rationality is to consist in the following:

People do what is the appropriate thing for them to do (i.e. people act rationally).	X wants E. X believes his circumstances to be C. If an agent wants E, then if the agent believes his circumstances to be C, the appropriate thing for that agent to do is A.

$\therefore$ X does A.

FOOTNOTES

[1]I am assuming, for the sake of case of exposition, that all other factors are irrelevant.

[2]The physical universe might be said to be rational in the sense that it can become an object of understanding for human beings. Cf. Michael Polanyi, Personal Knowledge (New York, 1958), ch. 1. But this is clearly a different sense of the term 'rational'.

[3]Wilfrid Sellars, for example, has discussed belief and inference in terms of such practical dimensions. Cf.: "Induction as Vindication", Philosophy of Science, vol. 31 (1964).

[4]For my purposes, "wanted ends" shall be taken to include also "the doing of duty for its own sake". Both action from desire and action from duty are thus to be regarded as rational in accord with my schema. Philosophers who raise certain moral problems by stressing the distinction include: Thomas Reid, Essays on the Active Powers of the Human Mind (Cambridge, 1969), Essay II, ch. 2; Essay III, Part III; Essay IV, ch. 4 (Reproduced from: The Works of Thomas Reid, vols III & IV, Charlestown, Mass., 1815); Immanuel Kant, Foundations of the Metaphysics of Morals, trans. by Lewis White Beck (Indianapolis, 1959), 2nd section; C. A. Campbell, On Selfhood and Godhood (London, 1957), pp. 148-156, and Lecture IX, passim.

[5]Two good introductions are the following: J. von Neumann and O. Morgenstern, Theory of Games and Economic Behaviour (Princeton, 1953); and R. D. Luce and H. Raiffa, Games and Decisions (New York, 1957).

[6]This point was suggested to me by Professor Robert W. Binkley.

[7]Cf. Objective Knowledge: An Evolutionary Approach (Oxford, 1972), especially "On the Theory of the Objective Mind", pp. 153-190, and "Of Clouds and Clocks", pp. 206-255.

[8]"Of Clouds and Clocks", p. 242.

[9]Ibid., p. 245.

[10]Ibid., pp. 246-248. This method of "error elimination" is exosomatic because it occurs with regard to evolved systems or organs which obtain outside the organism's body, for example, languages and theories, but which are nevertheless to be regarded as evolutionary advances of the organism, and hence of its species and phylum. See pp. 238-239.

[11]Ibid., pp. 249-250. A favoured example is that of the film of a soap bubble "plastically" or "softly" limiting the random motions of the air molecules enclosed within, yet itself partly sustained by those motions, for without them, the film of the bubble would collapse inward upon itself.

[12]It is, of course, the overall intent of this dissertation to support the view that rational explanation and its concomitant metaphysics of rationality are adequate to the subject matter of human behaviour as a whole.

[13]Cf. Laws and Explanation in History (Oxford, 1957), ch. 5.

[14]These pertinent cause-effect relationships would consist in those in which the "cause" was a possible action for the agent in the situation which he believes to obtain, and the "effect" was a wanted end.

VIII

PRINCIPLES OF ACTION

As amended, the schema for the explanation of human behaviour in terms of its rationality is the following:

People do what is the appropriate thing for them to do (i.e. people act rationally).	X wants E. X believes that he is in circumstances C. If an agent wants E, then if that agent believes that he is in circumstances C, the appropriate thing for that agent to do is A.

∴ X does A.

In order to establish the epistemological adequacy of this explanatory schema, I shall, in the present chapter, provide a defense of the empirical confirmability of general law statements of the sort shown on the right-hand side of the schema, hopefully demonstrating that the "evaluative" component expressed by the use of the term 'appropriate' is no bar to this confirmability. In the next chapter, I shall provide a defense of the conceptual adequacy of the framework principle, shown on the left-hand side of the schema, which, in effect, sets the pattern to which

explanations of human behaviour must conform in order that human behaviour be understood in terms of its rationality.

I

The first thing that must be noted about general laws of the form:

> If an agent wants E, then if that agent believes that he's in circumstances C, the appropriate thing for that agent to do is A.

is that there is nothing quite so elaborate as this in the common-sense fund of generalities that play an explicit role in understanding behaviour. It would normally seem rather contrived and verbose to make such statements as

> If a person wants to stay dry, then if that person believes that he's outdoors unsheltered in the rain, the appropriate thing for that agent to do is to seek shelter,
>
> If a person wants to eat, then if that person believes that he is in the kitchen with a bowl of beaten eggs and a frying pan, the appropriate thing for that person to do is to make an omelette.

Even if we suppose that reference to the agent's relevant wants may be implicitly assumed without having to be explicitly recited, it would still seem rather cumbersome, ordinarily, to phrase such locutions as

> If a person believes that he's outdoors unsheltered in the rain, then the appropriate thing for him to do is to seek shelter.

One reason for this persistent awkwardness is that ordinary practical generalities, or "principles of action", are usually phrased in terms of actual circumstances as judged to obtain by the speaker, and not in terms of circumstances the agent believes to obtain (as judged, of course, by the speaker). This difference has already been noted, and grounds have been presented for regarding the above schema to be ideally correct even if, in practice, not explicitly articulated in the usual commonsense explanatory remarks regarding behaviour. After all, it is even ordinarily acknowledged, if and when the question arises, that an agent can hardly be expected to have acted in a way which is appropriate relative to his circumstances if he was not aware that those circumstances obtained. It is simply that the agent's ability to be correctly aware of his circumstances--as correctly as we can determine them to be--is normally taken as the standard case, one which is to be expected. Thus only deviations from the normal case require special and explicit qualifications upon the explanations of the corresponding behaviour.

Another modification paralleling the above must be made in the schema of principles of action. If an agent drastically misconceives his circumstances then he may

believe himself able to accomplish something which he cannot in fact accomplish. It cannot *really* be appropriate for him to do something which he cannot in fact do, no matter how firmly he believes himself able to do that something. The most that can be said is that he should *endeavour* or *undertake* to perform the action which, from his point of view, appears potentially achievable. Given that his behaviour will be unsuccessful, it cannot ultimately be his doing of the action that we shall be explaining, for he will not accomplish what he endeavours to accomplish. The only proper way to understand *why* he did what he did, in terms of its rationality, is to grasp that action as the "attempt to do" the action which the agent *intended* to perform.

If the agent has correctly conceived his circumstances, then we still do not collapse the distinction between the actual circumstances and the agent's beliefs about his circumstances for the purpose of explanation. There is still a distinction between the agent's intended action and the agent's actual endeavour, with the difference that in this case the endeavour was successful, and hence the intended action did result. It is thus still only the *endeavour* which is explained by reference to the agent's wants and beliefs; the actual action must ideally be explained by reference to the *fact of endeavour* conjoined with the facts about *actual* (favourable) *circumstances*.

The explanation of actual behaviour, even in the cases of successful execution of intention in the manner intended, thus devolves into a two-stage process of reasoning, only the first part of which is really an explanation in terms of rationality, thus:

(Part I)

People endeavour to do what is the appropriate thing for them to do (i.e. people endeavour to act rationally).	X wants E. X believes that he is in circumstances C. If an agent wants E, then if that agent believes that he is in circumstances C, the appropriate thing for that agent to endeavour to do is A.

∴ X endeavours to do A.

(Part II)

X is in circumstances C.
X endeavours to do A.
If an agent is in circumstances C, then if that agent endeavours to do A, that agent does A.

∴ X does A.

It is clear that most ordinary explanations of successful action short-circuit what seems to be the ideal account by assuming the agent to be knowledgeable and aware until proven otherwise. Explanation is nearly always of behaviour rather than of endeavour, especially when this is

successful behaviour. Yet the endeavour can only be successful if circumstances permit. The idealized account of human behaviour must therefore reflect this distinction and ordinary explanations can only be adequate if they can, in principle, be supplemented with justified assumptions about the correctness of the agent's beliefs about his circumstances.

II

If the explanation of human behaviour were to rely upon laws of the form:

> If an agent wants E, then if that agent <u>is</u> <u>in</u> circumstances C, the appropriate thing for that agent to do is A,

rather than upon laws referring to the agent's <u>beliefs about</u> his circumstances, it might have seemed that behavioural explanation was relying upon laws which were peculiarly <u>evaluative</u> and in that respect <u>non-empirical</u>, in a way in which laws referring to the agent's beliefs about his circumstances are not. But careful inspection of the structure of the laws which are here phrased in terms of the agent's beliefs about his circumstances should reveal that no such difference exists. Both are "evaluative" in just the same crucial respect, namely the respect that they generalize regarding what is an "appropriate" thing to do

given certain conditions. For the laws which are utilized in the present context must be clearly distinguished from possible generalizations of the following sort:

> If an agent wants E, then if that agent believes that he is in circumstances C, that agent will endeavour to do A.

The laws which are utilized in the present context do not state the actual behavioural endeavour of the agent as a consequent of the agent's beliefs and wants, but rather state the appropriate behavioural endeavour for the agent as a consequent of the agent's beliefs and wants. If the judgment of appropriateness (of endeavour) is evaluative in any way which is peculiar and non-confirmable and which would therefore threaten the empirical testability of the rational form of explanation, then this threat to my account remains even though the laws involved pertain to the agent's beliefs about his circumstances. I shall attempt to show that claims about the appropriateness of action are as confirmable as any paradigm descriptive claims.

It can easily be seen that the following two forms of generalization, which we have succeeded in distinguishing from each other, correspond to each other in their constituent clauses:

(A)	(B)
(A1) If an agent wants E,	(B1) If an agent wants E,
(A2) then if that agent believes that he's in circumstances C,	(B2) then if that agent is in circumstances C
(A3) the appropriate thing for that agent to endeavour to do is A.	(B3) the appropriate thing for that agent to do is A.

Furthermore both of these forms of generalization bear a further correspondence to the following form of generality:

(C)

(C1) If E is wanted,
(C2) then if circumstances are of type C
(C3) the appropriate thing to do is A.

The correspondences seem to consist in the following relationships. Generality B is really a third-person prescriptive version of what generality C expresses in first person prescriptive, or "resolutive", form. Generality C is the form of the principle of action in use by an agent who must decide how he shall act, whereas generality (B) is a form of essentially the same principle but as it would be used by an observer of human behaviour who is deciding how to explain why the actions of someone else occurred. Generality (B) might be called a "translation" of generality (C) into the terms appropriate to a different standpoint regarding human behaviour. However, even a generality of form (B) carries a practical commitment: the explainer can see that he himself falls under the scope

of the prescription which the generality embodies, for he too can be an agent who must decide how he himself shall act. Nevertheless generality B is more explanatory than practical, or action-guiding, in its direction. It carries the burden of explanatory force just in those contexts where the assumption is made that the agents whose behaviour is being explained, conceive their situations as accurately as we the observers can determine them to be.

Generality (A), on the other hand, is formulated to take account of the difference between actual circumstances and the agent's beliefs about actual circumstances, and is almost fully explanatory, rather than practical, in its direction. An agent, in deciding how to act, does not ordinarily take a third-person standpoint to compare his own beliefs about his circumstances to some independent assessment of those circumstances; his beliefs about his circumstances constitute the only assessment of circumstances that he is able to make. Generality (A) reflects the standpoint of an observer of a behaving agent, and thus someone for whom the agent's beliefs and the circumstances can be assessed independently. From the standpoint of this distinction behaviour is rational so long as it is an endeavour which is appropriate to the agent's beliefs about his actual circumstances. Generality (A) thus takes into account the distinctions which only a third-person observer is able to make; hence it is a fully explanatory

formulation of a principle of action rather than a practical formulation of the same underlying principle.

The aspects of generality (A) which reflect the third-person observational standpoint pose no special problems for rational explanation. So long as the basic underlying prescription, namely generality (C) is itself confirmable, then generalities (A) and (B) are equally confirmable from their respective standpoints. Whatever special problem of evaluation, or prescription, versus description is to be found in the explanation of human behaviour in terms of its rationality, will be most clearly revealed in generality (C) which is the practical (and essential) core of principles of action. It is therefore to a consideration of the empirical meaningfulness of generalities of type (C) that we must now turn.

There are various ways of reformulating generality (C) by altering clause (C3), any of which may be suitable depending upon the context. Thus we might have

> If E is wanted, then if
> circumstances are of type C, ...
>
> ... the appropriate thing to do is A.
> ... A ought to be done.
> ... A is the thing to do
> ... A should be done,

and perhaps others. Any subtle differences between these formulations will be ignored in the present context, and

they shall be treated as if they were fully interchangeable salva veritate. The members of this family of hypothetical statements have been called "prescriptive" or "evaluative" by way of a presumed contrast with straightforwardly indicative or "descriptive" statements.[1] The legitimacy of this distinction is taken for granted by, for example, William Dray in his defense of a form of rational explanation which remotely resembles mine.[2] Dray's principles of action are of the form

> When in a situation of type $C_1 \ldots C_n$,
> the thing to do is X.

Dray contends that the infinitive "to do" in such principles functions as a value term thereby implicitly incorporating an element of appraisal into explanations which rely upon such generalizations.[3]

The generality or universality which Dray attributes to principles of action is thus conceived as being of a radically different sort than the generality of genuine empirical laws, a conception not peculiar to Dray alone. But this conception is frequently used to support a position according to which the intellectual adequacy of rational explanations is denied.[4] By Dray's own admission, people frequently do not actually do that which they ought to do; hence there is at least a prima facie plausibility in supposing that Dray's principles of action are not

empirical generalizations, but are simply evaluative. Left unqualified, this supposition would entail a serious weakness in Dray's position. James Leach, for example, confronts Dray with the presumed inadequacy of principles of action[5] on grounds which may be characterized in the following way: if conceived as being descriptive, then principles of action fail to be explanatory by virtue of failing to be universal; yet if conceived as being prescriptive, then principles of action fail to be epistemologically adequate by virtue of failing to be subject to empirical control.

Since the time when Hume disputed the possibility of deriving an 'ought' from an 'is',[6] it has been a common assumption that generalizations which formulate the conditions under which specified actions ought to occur are essentially non-empirical, and are, instead, of an epistemologically inferior status which may be called "prescriptive", "evaluative", or the like. It is only currently coming to be acknowledged[7] that prescription, rather than mutually excluding description, is simply one variety of it. The way things "ought to be" is as much a part of the way the world is as the way things actually occur. It indeed seems illegitimate to derive "X ought to be done" from "X is done", where 'ought' is understood in an action-guiding sense, and the whole derivation has the force of a commitment to some such generality as that "Actions which are done ought to be

done". Hume's injunction, understood in this eminently reasonable fashion is indisputable. But it is not so eminently reasonable to suppose that the derivation of "X ought to be done" from, say, "Y is the case" will also be clearly illegitimate in all cases.

The variously formulated principles of action may be said to comprise an "action-guiding" use of language. This use of language is based on the fact--or the presumption--that human action can be goal-directed, i.e. it can be performed for the sake of changes wrought in the world by so acting. Thus the proper use of principles of action implicitly acknowledges the wants and needs of human beings. As formulated in the present context, they are always hypothetical in construction, and are thus hypothetical prescriptions or imperatives. The 'ought' clause obtains conditionally, given some type of human want and certain types of circumstances. In order that these hypotheticals be empirically confirmable, it must therefore be possible to grasp the 'ought' clause as somehow following from the clauses expressing the want and the circumstances. An 'ought' must thus be derivable from an is, in accord with something like the following argument form:

E is wanted.
Circumstances are of type C.
∴ A ought to be done.

The empirical nature of principles of action can best be understood by first borrowing, from C. A. Mace and R. B. Braithwaite, certain concepts developed in regard to teleological processes and teleological explanations in general.[8] Mace conceives of a teleological process as a sequence of events characterized in the following ways: (1) it is evoked by the absence of a certain condition, and is of such a nature that the introduction of this condition at any time during the sequence would terminate the process; and (2) it is such that, through repeated occurrences, constituent events conducive to the attainment of the condition become stabilized, and constituent events adverse to the attainment of the condition become eliminated. As a result the process as a whole tends to achieve a unified form in which component events occur in a certain temporal order sufficient to produce the absent condition, each component being necessary within that process as a whole in order that the absent condition be produced by the process. This concept of teleological process clearly presupposes the notion of causality. It is most assuredly this notion which Mace has in mind when he points out that "A teleological sequence is a more or less economical arrangement of causes in a manner which secures the realization of an end".[9] The important point is that otherwise mechanically generated sequences of events are construed as admitting of conceptual organization in teleological form.

Similar notions of teleological process are suggested by Braithwaite's construal of teleological explanations as explanations which exhibit events in a system as being "nomically determined" by later events of a particular type in the same system. The determining event operates as a goal or end-state, such that the determined event initiates (or constitutes one link of) a causal chain of events which will culminate in the goal even under a (perhaps wide) variety of external conditions. Thus a teleological explanation of an event is not being construed as an explanation according to which the occurrence of the explanandum event is inferred on the basis of the occurrence of some antecedent event construed as causally efficacious. Rather, a teleological explanation is construed as an explanation according to which the occurrence of the explanandum event is inferred on the basis of some subsequent event or state toward the actualization of which the explanandum event can be expected to lead (under any of a range of favourable circumstances). Although Braithwaite only attributes value to teleological explanations in cases where non-teleological explanations are not available, nevertheless, in those cases, at any rate, he accepts teleological laws as legitimate conceptual devices by which we can organize our empirical knowledge and predict the future. On these grounds they are depicted as being no less intellectually satisfying, within the limits of their

proper application, than mechanistic laws exhibiting the physico-chemical determinancy of events on the basis of antecedent conditions.

The concepts of agency and of action for the sake of a wanted end (i.e. rational action) should facilitate understanding of the way in which principles of action are empirical generalizations which embody the Mace-Braithwaite concept of teleological process as "arrangement of causes in a manner which secures the realization of an end". For principles of action exhibit the (rational) necessity of performing certain actions (in particular circumstances) given the commitment to some goal. I hope to show that they are "translations" into first person practical action-guiding terms of certain ordinary causal generalizations of the following form:

If C obtains, then if A occurs, then E occurs.

The particular causal generalizations of this form which are of relevance in the present context are those in which 'C' pertains to circumstances which permit human action, 'A' pertains to an event which can be enacted, that is, it can constitute a human action, and 'E' pertains to a consequence of human action that might be wanted. A version of the above form of generalization which takes these distinctions into account would be the following:

If C obtains, then if A is done, then
E ensues (or is achieved).

Such generalizations are similar to the kind of generalization to which Wilfrid Sellars devotes special attention, in "Counterfactuals, Dispositions, and the Causal Modalities",[10] and which might be said to formulate, for scientific contexts, the practical, or action-guiding, import of ordinary empirical generalizations. These generalizations, like those of the previous sort, are clearly and straightforwardly empirical, a feature in no way altered by their formulation in terms which restrict the events under consideration to only those of a certain type, namely, human actions. The problem of their confirmation is the problem of the confirmation of empirical laws in general, and is no less resolved nor resolvable than the latter.

III

At this point a certain digression is necessary, and will eventually prove suggestive in a way which is fruitful for the present discussion. John Searle has argued[11] that a rigid dichotomy between descriptive and evaluative statements fails to give a coherent account of institutional, as distinct from brute, facts. By "institutional facts" Searle is referring to forms of human activity, specifically of a social nature, whose occurrence

proceeds in accordance with established social conventions, or institutions, to which human beings are committed. It is by means of such institutions for personal commitment, that, through social action itself, many obligations, rights, and responsibilities are created. Within the context of commitment to such institutions, certain otherwise brute facts of the world, for example certain bodily movements, can constitute the creation of some specific obligation towards a specific individual, as in, for example, promising or marrying. And although Searle does not explicitly say this, it seems clearly true that many otherwise brute facts of the world, which by institutionalized convention constitute institutionally significant facts, would simply not occur even as <u>brute</u> facts in the absence of institutions which impart to them this special significance. Institutional facts are not just brute facts seen through blinkers of commitment, and praised or decried on the basis of that commitment. Institutional facts are institutionally significant brute facts about human interrelationships the brute <u>occurrence</u> of which only obtains in <u>virtue</u> of the existence of institutions which prescribe that they <u>constitute</u> peculiarly committed actions. There simply would be no such brute facts as the utterance, in any language, of the statement "I promise to pay you, Jones, five dollars for mowing my lawn", were it not for the institution of promise-keeping.

Searle's concept of "institutional facts" seems peculiarly relevant to the moral use of "ought". I would, therefore, like to propose an analogous concept which pertains to "ought" discourse in general, and which I call the concept of "personal facts". The important notion behind this proposal is that of "personal commitment", especially one's commitment to oneself to survive and to live a full and satisfactory life. Within the context of this all-encompassing personal commitment occur the more specific and immediate commitments of, in Kantian terms,[12] willing the indispensably necessary means to one's ends, that lie within one's power. That is, the commitment to personal survival and welfare encompasses and enjoins all the commitments to satisfy more immediate wants and needs, by effecting a constraint upon the will which renders (rationally) necessary, for the agent, the performance of some action toward the immediate goals. This is almost to make a definitive point about the concept of wanting, namely, that it involves a particular kind of mental "set", an inclining of the mind toward a certain achievement, which inclining generally manifests itself as the overt behaviour of striving toward that achievement as circumstances permit.[13] Of course such overt behaviour need not necessarily occur in order that it still be true of some person that he has some particular want. But in order that personal human behaviour, in general, not constitute merely

random and incidental movements of merely living organisms, it must in general be selectively contrived and directed by those active organisms themselves. And such direction only makes sense as the natural output of personal wants and such personal needs as are acknowledged as such by the agent.

Within the context of personal commitments, otherwise brute facts take on a new significance. The otherwise brute facts of mental states of wanting constitute personally significant facts from which it can be inferred that something ought to be done by way of attempting to achieve the appropriate personal satisfactions. And where more specific circumstances make successful activity in this direction possible, then the otherwise brute facts of the obtainment of a particular want and particular sorts of circumstances can analogously be the basis for inferring that some specific sort of action on the part of the agent is then (rationally) obligatory. Furthermore otherwise brute facts of the natural, non-human world, can constitute personally significant facts about things which can be wanted, things which can be achieved, and situations in which wants and needs can be satisfied. Finally, as in the case of Searle's concept of institutionally committed behaviour, those brute facts which constitute personally directed and committed behaviour are not mere brute facts seen in distorted perspective; they would not even occur as brute facts were it not for the human wants and needs which

they, in general, satisfy and for the sake of which they are performed.

It is a mistake to think that the contrast between brute and personal facts marks a contrast between non-evaluated and evaluated facts. For if we suppose that brute facts are such as would be described ultimately in pure sense terms, then there is a clear sense in which even brute facts are not described from a neutral impersonal, hence non-evaluative, point of view. I do not have in mind the uninteresting point that the use of any term for the purposes of description involves an evaluation of the appropriateness of that term to the situation being described. I have in mind, rather, the sort of thesis put forward by, for example, Stuart Hampshire in Thought and Action.[14] Hampshire argues essentially that the notion of a thinking observer who is not also an active being participating in and among the surroundings which he observes is incoherent. Perception only takes the form that we know it to take by essentially involving self-consciousness, the sense of self as observer and of external objects as the observed, and also by virtue of the fact that the perceptual discriminatory abilities of the observer depend upon his ability to manipulate, to be an active being in regard to that which is observed.[15]

For Hampshire, the very forms of language in which we describe the world are determined by our practical

interests. We can, of course, imagine observation to occur from a point of view for which, as Hampshire expresses it, thought of the possibility of action, based upon what is observed to occur, is temporarily suspended. However this suspension does not eliminate the way in which the sensory and linguistic capacities which we should transfer to our quasi-neutral observational standpoint will still reflect, through their nature, our usual ongoing situation as interested observing participants in the observed world. At the very least, the data of pure sense are determined by the structures and limitations of our sense capacities. And this is to say that even brute facts are "personal" facts, constituting the world as described from the point of view of a being with human sense and intellectual capacities, i.e., personally significant sense and intellectual capacities.

Furthermore, it is not altogether obvious that what we conceive as a most "neutral" point of view is actually neutral in any neutral sense, i.e., as itself evaluated from any further neutral or non-humanly-interested standpoint. We can only establish its "neutrality" from our personal standpoint. It does not seem that a set of wants and needs combined with a preparedness to act will necessarily corrupt the observations made by someone characterized by those conditions. After all, in order that actions be efficacious in fulfilling their intended purposes,

it is necessary, in general, that the observations upon which those actions are based be largely correct. Similarly, it does not seem that observations which are made by someone for whom participation and selective interest are suspended will necessarily lack corruption. Thus, insofar as "neutral" is construed as a synonym for "uncorrupted", there appears to be no strong argument that an interested standpoint is necessarily not also neutral. Finally, as Michael Polanyi argues, in Personal Knowledge, even a quasi-neutral scientific point of view reflects a very highly personal aesthetic sense, our intellectual pleasure in the apprehension of such rational factors in the natural world as mathematical order and harmony.[16] We may wish to distinguish the scientific point of view from other (humanly possible) points of view as one which is distinctively characterized by this (personal) aesthetic sense; but we have not thereby distinguished a point of view which is essentially "neutral" or "value-free" from those which are essentially not.

If it is not insisted upon that brute facts are such as must be described in pure sense terms, but if it be allowed that our ordinary middle-sized object language can be used to describe the world in a brute, or "neutral", fashion, then the case becomes even stronger for denying the existence of any essential epistemological difference between such so-called "neutral" descriptions and the uses

of "ought" discourse. As Julius Kovesi has pointed out,[17] between such extreme divergent concepts as 'good' and 'yellow', lie a whole world of further concepts which, like that of 'table', are not peculiarly part of moral discourse, yet which "supervene", in a manner usually only attributed to peculiarly moral terms, the sums of the sense properties of their instances. Our understanding and use of such terms cannot be exhaustively represented by extensional listings of what Kovesi calls the "material elements" of each of their instantiations. Tables are objects which human beings create in order to use, and use in order to satisfy certain needs in their lives. This distinctive usefulness of actual tables is what Kovesi calls the "formal element", the only essential element in the concept of a table; it comprises the point of taking the many material elements of any one particular material object to be a table. This "formal element" is not itself grasped extensionally either, for there is no necessary limit to the variety of qualities which an object can have yet still be a table. Of course, in a certain sense, there is not anything which is "table" over and above all the particular material objects in space and time which were, are, and will be, tables. But at any given time during the natural history of a living language, all future and novel instances of then current concepts are simply not yet known. Kovesi's "formal element" is akin to what Sellars calls the

"continuing theme" of a term, which guides it through the "vicissitudes of empirical knowledge".[18] Because of the way in which tables are created to suit human wants and needs, human values play an essential role in the ongoing descriptive application of the concept of "table" to the world.

IV

With the above point in mind, we can return to the main discussion of this chapter. Consider now the practical significance of the fact that, given certain sorts of circumstances C', the performance of one action of the type A' will bring about the achievement of an end-state of type E'. It is, I maintain, a personal "practical truism", that in such circumstances, an action of type A' is an appropriate thing to do in order to achieve an end-state of type E'. I am suggesting the practical equivalence of the following two empirical generalizations:

> If C' obtains, then if A' is done, then E' is achieved,

and

> If C' obtains, then if the achievement of E' is wanted, then A' is an appropriate thing to do.

Of course, it cannot necessarily be said that A' is the appropriate thing to do, since, in the given circumstances,

alternative actions may bring about achievement of the same (wanted) end-state. Therefore, in the absence of further information, an action of type A' can be considered as only one among a possible plurality of appropriate things to do.

In any case, the above "practical truism" reflects two important <u>personal</u> facts about agents in the world: first, the end-states are not merely effects which incidentally occur as a consequence of behaviour, but are effects which can be <u>wanted</u> (in the loosest sense of this term); and, second, that behaviour need not give rise to wanted effects in a merely incidental way, but may specifically and essentially be enacted <u>for the sake of</u> giving rise to such effects. As Robert Binkley has pointed out,[19] means-end reasoning does not merely exhibit the interconnections of things, but embodies (the reasoner's) self-consciousness of personal effectiveness as an agent causing things to happen; and, I might add, as an agent <u>for whom</u> things happen.

It is crucial to recognize that empirical generalizations of the following form:

> If C obtains, then if the achievement of E is wanted, then A is an appropriate thing to do

involve no (additional) element of appraisal or evaluation over and above the very empirical regularity itself as seen in terms of its personal practical import. And although it

may be true to say that the terms of personal practical import do indeed place the empirical regularity within a context of "values", nevertheless this is only to say that evaluation, appraisal, prescription, and the like--as exemplified above, constitute nothing more than the description of the empirical regularities of the world as they pertain to the behaviour of a being for whom otherwise "neutral" events may be wanted as end-states, and who is capable of controlling the chains of actual events in the world to the extent of ultimately causing the occurrence of such end-states.

A concept of "practical knowing" has been recently developed according to which a person's knowledge of his own body and of his behaviour is thought to be different in kind from the knowledge gained by mere (passive) observation.[20] Practical knowing, on this conception, essentially reflects the fact that the knower is an agent in action. This concept, as suggested in the previous section, can be extended beyond simply knowledge of one's own body and one's own behaviour to encompass a personal practical knowledge of the world, a kind of conversion of the (non-practical) knowledge gained by mere observation into a form which has an immediate relevance to action, and especially, to one's own activity. Thus it is that prescription is simply description from the point of view of an agent who is active and effective in regard to that which is described.

But I have not yet succeeded in defending the form of prescriptive statement exhibited by principles of action. Further conceptual manipulation is required. Consider the fact that, in given circumstances, if a variety of actions will result in achievement of a particular type of end-state, then, other things being equal, each such action is an appropriate thing to do in order to achieve that end-state. On these grounds alone, it is not possible to select a <u>particular</u> action as being <u>the</u> appropriate thing to do. However, it seems clearly possible to say that in the given circumstances <u>one</u> of the actions which will achieve a certain end-state is <u>the</u> appropriate thing to do in <u>order</u> to achieve that end-state. This reflects another "practical truism" that where an end-state may be achieved in a variety of ways which, as a group, are such that no rational choice may be made among them, nevertheless, the want of the specified end-state rationally determines that <u>some</u> choice among them, however arbitrary, be made.

The relevant principle of action in such a case would involve a general enough specification of the type of action appropriate to perform under the given conditions so that this type-description would encompass all the different ways of achieving the wanted end among which no rational difference exists. For example suppose that one wants an economical fuel-less means of transportation for use in a small town, and suppose that one is in good health and has

$100 to spend. The appropriate thing to do would be to buy a bicycle. But there may be hundreds among which one could choose. If no further factors matter to the agent regarding which particular bicycle it shall be, then the relevant principle of action need merely describe "the appropriate thing to do" at a level of generality which is such as to be satisfied by the act of choosing any one of the bicycles in the class of equally rational purchases, thus ". . . the appropriate thing to do is to buy a (any) bicycle costing under $100".

The above sorts of principles of action contain no special or distinctive elements that cannot be understood as having empirical significance. On these grounds, such generalizations should be found to admit of empirical test. And clearly this is the case. A principle of action is tested whenever the type of action which it enjoins in a particular type of situation, given certain wants, is performed in such a situation. To the extent to which the action does bring about an end-state which is what was wanted, then to that extent the principle of action is confirmed in any single instance of its application. Counter-instances to such generalizations consist in the failure of a wanted end-state to be achieved following performance of the action which was judged to be appropriate for its attainment. In such cases, it is clearly a matter of empirical fact that in the given situation, the action

actually performed was not the appropriate action for that agent to have performed; i.e., that agent ought not to have done what he did.

Even if a large number of people fail to act in accordance with some particular principle of action, its generality and concomitant explanatory relevance are not thereby impugned. For some particular action may indeed be the thing for people, given certain wants on their part, to do in certain circumstances, even if few people ever actually succeed in doing it. People may fail to act in accordance with a given principle of action if another want assumes a greater importance in some situation in which action to achieve either excludes action to achieve the other. As noted previously, a human being is characterized by a hierarchy of wants, several of which may seem achievable in a given situation. Each principle of action relevant to the situation only formulates what is appropriate for a given agent to do relative to certain selected conditions, a certain want or wants, and certain situational circumstances. The reconciliation of all the principles of action that are relevant to a given agent in a given situation provides that agent with a "decision problem" and may require a process of judgment for which no further generalization may be available. Thus few people may ever succeed in doing action A' in circumstances C' even though it is appropriate to achieving end E' perhaps

because end E' is never very important to people in circumstances C': it might be a life and death situation in which end E' is nearly always superceded in importance by, say, end E'', the desire to stay alive. This does not mean that the relevant principle of action is in any wrong, but only that it may have limited practical importance.

V

I have been maintaining that the 'ought' clause of principles of action must be thought to follow from the two antecedent clauses, pertaining to wants and circumstances, if the entire form of generality is to constitute an empirically meaningful and testable whole. Max Black has defended[21] a similar form of derivation, namely:

Jones wants to x.
The one and only way to x is for Jones to y.
∴ Jones should y.

Although I am in complete agreement with Black's argument, he does not seem to me to have fortified his position with sufficient qualifications as to defend it against a rather obvious line of attack. The previous remarks regarding the possibility of a practical "decision-problem" for an agent should enable such fortification.

The obvious line of attack is that often that which one wants to do is something which one should not do (ought not to do, is not appropriate for one to do). The

appropriate response, in defense of Black's (and my) form of derivation, is that the conclusion of a derivation is dependent on its premises and may only be true conditionally in respect of those premises. Thus it is only in so far as Jones wants to x and the one and only way to x is for Jones to y, that Jones should y. To the extent that other factors of the situation entail that Jones should not y, then in so far as those other factors obtain, Jones should not y.

The point being suggested with regard to the appropriateness, or rationality, of action parallels that made by Sir David Ross regarding the nature of moral duty.[22] Ross uses the term "prima facie duty" to refer to a characteristic

> . . . which an act has, in virtue of being of a certain kind (e.g. the keeping of a promise), of being an act which would be a duty proper if it were not at the same time of another kind which is morally significant.[23]

This characteristic

> . . . is an objective fact involved in the nature of the situation, or more strictly in an element of its nature, though not, as duty proper does, arising from its whole nature.[24]

Prima facie duties are distinguished by Ross from what he calls actual or "absolute" duties, and which are characteristics of acts in virtue of their whole natures, not

just some one component of their natures as was the case with regard to prima facie duties. Ross regards the prima facie rightness of certain types of act as being self-evident to any one of sufficient maturity who gives sufficient attention to the matter--self-evident in the manner of a mathematical axiom or the validity of a form of inference. However he regards our judgments about our actual duties in concrete situations as having none of this certainty. In an actual situation, where we apprehend two or more prima facie duties as being relevant, then we may not be certain whether we ought or ought not do the thing, in an absolute sense of "ought", and, as Ross puts it, ". . . whether we do it or not, we are taking a moral risk".[25] Even if we only apprehend one prima facie duty as being relevant in a given case, we cannot be certain that there are not further morally relevant factors which have been overlooked. There is no principle by which we can finally draw the conclusion that an act is on the whole right or on the whole wrong. Our final considered judgments in such cases are ". . . more or less probable opinions which are not logically justified conclusions from the general principles that are recognized as self-evident".[26]

Assuming the correctness of Ross' distinction, it follows that a given action can be both morally right and morally wrong for some person in a given situation, so long

as these assessments are not both to be understood as the assessments of actual, or "absolute", duty. At least one, and possibly both, must be only assessments of prima facie duty. But if this is the case then no morally relevant contradiction need arise, for the act can clearly be morally right in one respect while morally wrong in another.

The soundness of Ross' moral views are not of issue here; what is useful, however is the notion of prima facie versus absolute duties, and of prima facie duties in respect of certain (isolated) characteristics of a situation. There is no reason why a certain action for a given agent in a given situation should not be appropriate, or rational, in one respect and inappropriate, or irrational, in another. This possibility derives from the fact that a given agent at a given time may be characterized by a hierarchy of standing wants, and a given situation may render the attainment of several of these possible though not necessarily jointly possible. Furthermore, some of these actions may bring about consequences which the agent wants to avoid, for example that of personal injury. Given such "decision problems" for the agent, what shall the agent actually do?

This question can be taken in at least two different senses. If the question is: "In general, what shall such a person do?", the answer can only consist of the obvious, but, in a practical sense, unhelpful, formula

that, in general, such a person shall act so as to achieve the most preferable state of affairs given the whole set of his needs and wants. If the question is: "What shall such a person, in this or that specific case, do?", then it is only possible to give an answer in situ, where the actual circumstances may be confronted and grasped in their entirety. It might be possible to work out a general policy for cases of conflict between certain specific types of wants. For example, in political affairs, it is frequently necessary to sacrifice general political liberty in the interests of economic advancement, or vice versa, and some further factor might be imposed in order to adjudicate between these basic interests. However such policies of compromise are still not sufficiently particular to determine what a particular agent shall do in a specific situation for the reason that the stated policy for conflict resolution does not encompass all the possible wants and needs of an individual which might, in an actual situation, intervene to override the obligation of the individual to act in accordance with the stated policy.

There seems to be one kind of feature of situations which we can never be rationally justified in believing, namely, that some determinate account of the elements of that situation does indeed exhaust all the elements of that situation. This would be a "fact" expressible by saying ". . . and there is nothing else".

Only if it is known that there is no factor in a situation which remains to be accounted for can it be known that nothing has been ignored which is relevant to a practical recommendation based upon that situation. But like Descartes' evil demon, the possibility of the existence of something else, as yet overlooked, simply cannot be exorcised by any rational procedure from our grasp of actual situations in the world. Of course, as with the evil demon again, there may be no rational ground for supposing that there _is_ some relevant factor in a situation which has been overlooked. But this observation does not eliminate the epistemological problem.

We may, furthermore, take note of situational factors to which the current state of action-guiding knowledge does not direct our attention at all, yet which we nevertheless do not entirely dismiss as irrelevant when we assess those situations. It is always, in principle, possible that the current state of action-guiding knowledge does not legitimize belief in the truth of something which in fact happens to be true. Just because our theoretical knowledge is far from complete, and may, furthermore, not be knowledge at all but false belief, therefore the respects in which an actual or "absolute" practical "ought" statement is not justifiable by reference to currently systematized _prima facie_ principles of action are not necessarily the respects in which such a statement is, in

an absolute sense, unjustified. Indeed new understanding can only be acquired through the process of coming to believe that relationships exist which were not before recognized, or that situational factors have action-guiding relevance which was not formerly recognized as such.

H. D. Aiken makes a similar point in reference to moral principles.[27] Aiken notes that because "No action is exhaustively definable simply as an act of a certain sort . . .", therefore an act may exhibit morally relevant aspects that have not yet been covered by prevailing moral principles. It may therefore be maintained that some action ought to be performed even though it ". . . goes against every rule in the book". Of course, to be reasonable, such a judgment must acquire support upon a ". . . full and impartial review of the whole moral situation . . ."; it cannot continue to be dogmatically maintained even after such a review uncovers no support for it. Such a judgment may therefore leave generalizability in its wake, in the form of a new moral principle. There seems to be no reason why actions may not be similarly appropriate to achieving certain wants in given situations in respect of characteristics that have not as yet been recognized as such. Consequently the judgment that someone should, in an absolute sense, act in a certain way at a certain time may fly in the face of all known principles of action, yet (later) prove to have been a correct assessment after all.

But if this is the case, it should be because the "proof" will have revealed some new principle of action which was unrecognized at the time yet is later seen to have had overriding importance in the situation in question.

It is a frequent argument[28] that "ought" statements (unqualified) cannot be general or law-like in the same way as scientific laws of nature because, in their usual abbreviated formulations, the former clearly admit of not infrequent exceptions, yet in sufficiently extended and qualified formulations, they lose their generality by coming to describe highly specific types of situations, until ultimately they only characterize particular cases. This mistaken argument apparently rests upon a failure to distinguish the role of prima facie, or general, principles of action from that of absolute, particular "ought" statements. Prima facie "ought" principles are not intended as exceptionless and merely general descriptions of a number of actual situations, but as contrived abstractions which express in idealized form, relationships among various selected factors, which are necessary relationships precisely when, and only when, regarded independently of the potential interfering conditions of actual situations.

But as Ross points out, this appears to be no less than what may be said about scientific laws of nature.[29] Ross gives the example of an object which, because it is subject to the force of gravitation of some body, say, the

earth, will tend to move in a particular direction with a particular velocity. But in so far as it is also, at the same time, subject to the action of other forces, it will diverge from the behaviour which would result if it were merely subject to the force of the earth's gravitation. Yet the law of gravitation is nonetheless true for this divergence. Recognition of this distinction is, according to Ross, what enables us to preserve the "absoluteness of laws of nature". To refer to a poor but not unfamiliar relation of laws of nature as an example, when we state such familiarities as that matches light when scratched, we do not need to be told that this process obtains only among suitable circumstances; it so happens that circumstances are generally suitable, and we take <u>this</u> fact for granted. We, therefore, only make a point of noting <u>exceptional</u> circumstances, these being, as a matter of fact, infrequent, and hence, as a matter of intellectual economy, more readily encompassed, than the more frequent, suitable circumstances, as <u>explicit</u> factors in our grasp of the behaviour of matches which are struck. Yet even this grasp is open-ended, in that it must allow for the possibility that as yet unforeseen circumstances may sometime obtain upon the occasion on which a match is struck which prevent that match from lighting. The abbreviated expression of natural laws, though risking non-validity in this way, possess the virtue of facilitating our comprehension of them, while yet

providing a reasonably sound structural framework within which the most common situations can be explained handily and only the infrequent recalcitrant instances will require explicit appeal to exotic auxiliary explanatory principles.

As there appeared to be a logical gap between the prima facie, general and the absolute, particular "ought" statements, so there appears to be a similar logical gap between what may be called the prima facie, general and the absolute, particular sorts of "is" statement. If we attempt to explain or predict an actual occurrence or state of affairs, then our conclusion must somehow encompass the total variety of detail present in the actual situation, for to be a genuine explanation or prediction, the conclusion must apply to a certain particular situation, and must do so by implicitly acknowledging the individuality and uniqueness of that situation in its entirety. But scientific theories may entitle us to rationally justified expectations with regard to only limited and specifiable factors. An absolute, particular "is" statement assesses, by implication, all details of a situation; it is not merely a summary of the situation's theoretically relevant details.

A schematic illustration may make my point clearer. Suppose that the scientific theoretical apparatus implies the generalization that circumstances $C_1 \ldots C_n$ constitute a set of causally sufficient conditions for the occurrence

of event E. Upon the actual attainment of circumstances $C_1 \ldots C_n$ within some wider situation, we are rationally entitled to draw the obvious conclusion that E will occur. But shall this be our actual judgment with regard to the actual situation? Does our inference above not require the additional qualification that no <u>further</u> circumstances were present which <u>obstructed</u>, in some way, the sufficiency of $C_1 \ldots C_n$ to be attended by the occurrence of E —— requires this qualification if it is to be the correct assessment of the actual situation? Or, put another way round, does not a final judgment that E will occur <u>reflect</u>, by implication, <u>this very assessment</u> of the situation at hand? The truth of such a judgment depends no less upon the truth of this qualification than it does upon the truth of any of its contributing theoretical laws. Yet this qualification, like the non-existence of the Cartesian demon once again, is simply not demonstrable in any case so long as the corpus of human knowledge remains incomplete. I do not deny that the truth of this qualification can <u>obtain</u> in any case; I argue only that it is logically impossible that there be general principles underlying the <u>belief</u> in its obtainment in any case.

If it is true, as I would maintain, that we understand a situation by grasping it as part of a potentially recurrent pattern within some wider context, then it would seem to follow that even the "intuitive" grasp of a

situation must, to constitute genuine understanding, involve implicit generalizability in some way. By virtue of the meaning of 'intuition', such generalizability cannot consist in the awareness that previously acknowledged general beliefs exhaustively justify the case at hand. The implicit generalizability of intuitive understanding can only consist in a modification of the conceptual "set" which the knower, as a matter of course, brings to his appreciation of the world around him. Thus intuition, for "is" statements as well as "ought" statements, constitutes true understanding by leaving generalizability in its wake, through the initiation of a new inference pattern for future intellectual comprehension of the actual diversity of the world.

Nevertheless, the point remains that at the time of their occurrence our absolute, particular assessments of actual situations implicitly involve an element of pure appraisal, of pure evaluation. This point seems to be equally true of both "is" and "ought" statements. In this logical matter then, "is" and "ought" statements are indistinguishable one from the other. And if, as I have claimed, general "ought" statements are as logically and empirically respectable as are general "is" statements, i.e., scientific laws of nature, then the descriptive-evaluative distinction would seem to cut across the total sets of "is" and "ought" statements indiscriminately,

rather than distinguishing these two sorts from each other.

All of these considerations lead to the conclusion that it can be true that, in a *prima facie* sense, Jones should y even while it is true that Jones should *not* y, so long as this is either only *prima facie* true as well, but in virtue of different aspects of the total situation, or else absolutely true, which is to say that it is a final considered judgment presuming to summarize *all* the aspects of the situation that have been grasped as being action-relevant. Furthermore, in this respect, "ought" conclusions are not epistemologically different from "is" conclusions. Thus, there seems to be no reason why conclusions such as

. . . the appropriate thing to do is A

should not be derivable from premises of the form

E is wanted
Circumstances are of type C

so long as it may be assumed that action A will *achieve* end E *in* circumstances C. But *this* generalization, if added to the given premises, is certainly not unempirical or non-confirmable. The entire argument constitutes the derivation of an 'ought' from an 'is'. And the principle of action which is the hypothetical constructed out of the above premises and the conclusion of such a derivation is confirmable precisely in accord with the evidence for the generalization which was *assumed* in the derivation, namely:

> If circumstances C obtain,
> then if (action) A is done, then
> (end) E is achieved.

Finally, since first-person practical principles of action are confirmable, then the same must be said for the third-person explanatory principles of action which are merely "translations" of the former into terms reflecting a spectator's rather than an agent's point of view, thus:

> If an agent wants E, then if that
> agent believes that he's in
> circumstances C, the appropriate
> thing for that agent to endeavour
> to do is A.

FOOTNOTES

[1]Cf. R. M. Hare, The Language of Morals (New York, 1952), ch. 10.

[2]Cf. Laws and Explanation in History (Oxford, 1957), ch. 5.

[3]Ibid., p. 124.

[4]Cf. Morton White, Foundations of Historical Knowledge (New York, 1965), p. 183.

[5]"Dray on Rational Action", Philosophy of Science, vol. 33 (1966), pp. 63-65.

[6]Treatise on Human Nature, edited by L. A. Selby-Bigge (Oxford, 1888), Bk. III, Pt. I, pp. 469-470.

[7]Cf. Julius Kovesi, Moral Notions (New York, 1967).

[8]C. A. Mace, "Mechanical and Teleological Causation", in Herbert Feigl and Wilfrid Sellars, eds., Readings in Philosophical Analysis (New York, 1949), pp. 534-539. R. B. Braithwaite, Scientific Explanation (New York, 1953), esp. ch. 10.

[9]Op. cit., p. 538.

[10]In: Herbert Feigl, Michael Scriven, and Grover Maxwell, eds., Minnesota Studies in the Philosophy of Science, vol. II (Minneapolis, 1958), esp. parts I and II.

[11]Cf. "How to Derive an 'Ought' From an 'Is'", The Philosophical Review, vol. 73 (1964), pp. 43-58; also Speech Acts (Cambridge, 1969), esp. ch. 8.

[12]Foundations of the Metaphysics of Morals, trans. by Lewis White Beck (Indianapolis, 1959), pp. 64-65.

[13]A similar point is made by Max Black, "The Gap Between 'Is' and 'Should'", in W. D. Hudson, ed., The Is-Ought Question (London, 1969), p. 103; and by G. E. M. Anscombe, Intention (Oxford, 1957), p. 67.

[14](London, 1960).

[15]Ibid., pp. 47-51, 67-70.

[16](New York, 1958), cf. pp. 6-15.

[17]Moral Notions (New York, 1967), esp. ch. 1.

[18]Op. cit., p. 261.

[19]"A Theory of Practical Reasoning", The Philosophical Review, vol. 74 (1965), p. 443.

[20]Cf. Anscombe, op. cit., pp. 57, 77, 82-89; John Yolton, "Agent Causality", American Philosophical Quarterly, vol. 3 (1966), pp. 24-26.

[21]Op. cit., pp. 136-137.

[22]The Right and the Good (Oxford, 1930), pp. 19-20, 28-31.

[23]Ibid., p. 19.

[24]Ibid., p. 20.

[25]*Ibid.*, p. 30.

[26]*Ibid.*, p. 31.

[27]"The Concept of Moral Objectivity", in Hector-Neri Castaneda and George Nakhnikian, eds., *Morality and the Language of Conduct* (Detroit, 1965), pp. 99-100.

[28]Cf. Dray, *op. cit.*, pp. 28-44.

[29]*Op. cit.*, pp. 28-29. Cf. also Michael Scriven, "The Key Property of Physical Laws--Inaccuracy", in Herbert Feigl and Grover Maxwell, eds., *Current Issues in the Philosophy of Science* (New York, 1961), pp. 91-104.

IX

THE PRESUMPTION OF HUMAN RATIONALITY

In this chapter, I shall provide a defense of the conceptual adequacy of the framework principle involved in the explanation of human behaviour in terms of its rationality. This principle, in effect, sets the pattern to which explanation of human behaviour must conform if that behaviour is really to be understood in terms of its rationality, and not simply as mere regularity of undetermined sort. It may be recalled that this framework principle is as follows:

> People endeavour to do what is appropriate for them to do (i.e., people endeavour to act rationally).

This principle sets the pattern to which rational explanation must conform by, in effect, requiring that evidence for the <u>actual</u> performance of behaviour shall consist in data concerning the antecedents of behaviour which are such that the behaviour in question can be appreciated as <u>appropriate</u>, or <u>rational</u>. The framework principle is itself the presumption that this information shall naturally warrant the inference that the behaviour in question actually occurred (unless further <u>positive</u> evidence

indicates that the behavioural event did not occur as a result of the agent's appreciation of its appropriateness).

I shall defend this principle by way of a criticism of a form of explanatory schema for human behaviour which runs contrary to that presented here. This schema is introduced by Carl Hempel after certain criticisms are noted with regard to an earlier form of rational explanation remotely similar to my own which was introduced by William Dray.[1] This form of explanation may be schematized as follows:

> A was in a situation of type C.
> In a situation of type C, the appropriate thing to do is X.
> ---
> Therefore A did X.

The laws involved in this schema can be seen to be of the type (C) set forth in the preceding chapter, and which were noted to reflect a first-person practical viewpoint and to constitute the empirical core of the more idealized formulations of law which I employ, namely:

> If an agent wants E, then if that agent is in circumstances C, the appropriate thing for that agent to endeavour to do is A.

As a minor point, it may be noted that the distinction which I have belaboured between the wants and beliefs of an agent need not be thought of as having been disregarded by Dray's form of rational explanations as long as they can be

subsumed either under the definition of the situation, C, or the identification of the agent, A.

Carl Hempel and Alan Donagan have objected on a variety of grounds to Dray's form of rational explanation.[2] According to one such objection, the form of rational explanation which utilizes principles of action is defective by virtue of presupposing a criterion of rationality which has the unlikely effect that one particular course of action is singled out as being the thing to do in a given situation.[3] But this objection is not one which reveals any formal inadequacy in the form of rational explanation which relies on principles of action. It amounts, rather, to a (probably legitimate) claim that not all human behaviour can be exhaustively accounted for on the basis of laws which do single out one particular course of action as being the thing to do in a given situation. I have already argued that in cases where a variety of actions are equally appropriate in a situation, given the entire network of relevant wants, purposes, etc., then there can be no rational basis for the performance of any one particular action from among this variety in contrast to the performance of any other. Consequently, such a performance, in respect of its character as the manifestation of a choice among equally rational alternatives, cannot be understood or explained in terms of its rationality. But in respect of its character as the manifestation of a choice of some

one action from among the set of equally (most) rational alternatives, as contrasted with a choice of some action not at all of this set, such a performance is rational and to that extent, can be so understood and explained. The argument based upon the non-rationality (note: not irrationality) of some actions, or of some actions to some degree, highlights only the (uninteresting) fact that the form of rational explanation, and hence the understanding of behaviour in terms of its rationality, is limited in its application to behaviour which is, to some degree, genuinely rational, and only to the extent that it is thus rational.

More fundamental is the objection[4] that rational explanations which utilize principles of action simply do not explain why particular agents in fact committed the actions which those explanations purport to explain. In accord with the introductory remarks of the present chapter, this objection can be seen as a challenge to the possibility that conclusions regarding the actual performance of particular types of actions in particular types of situations, given certain wants, can in fact be derived from laws regarding the appropriateness of those same types of actions in those same types of sets of circumstances. As I have suggested, the legitimacy of such derivation rests upon the framework principle which amounts to a standing presupposition of the inherent (practical) rationality of human beings. Let us contrast this view with an attempt to

construct a schema for the explanation of human "rational" behaviour in the absence of such a guiding framework principle, and see whether or not such explanation could be satisfactory as the understanding of human behaviour in terms of its rationality.

Most notably, the major generality assessing the appropriateness of actions given their antecedents shall have to be replaced by a form of generality which is thought to lack the so-called evaluative aspect of the former (examined in the preceding chapter) and which merely expresses the occurrence of actions given their antecedents. Such a form of generalization gives rise to the impression that an action is a mere causal consequence of its antecedents. Thus, Hempel proposes to amend the form of rational explanation which utilizes principles of action by replacing those principles with so-called "descriptive statements" to the effect that the agent in question was a rational agent, or was disposed to acting rationally at the time of the action under consideration, and that a rational agent, when in a situation of the type in which the agent in question found himself, will always (or with high probability) do the sort of action for which the explanation is being given. Construed along Hempelian lines, a proper rational explanation would exhibit the following form:

A was in a situation of type C.
(H1) A was a rational agent.
(H2) In a situation of type C, any rational agent will do X.
Therefore A did X.[5]

Hempel defends his use of the substituted premises H1 and H2 by appealing to what he calls the "descriptive-psychological" sense of the notion of rationality. This is presumed to be a notion of rationality the application of which is governed by objective empirical rules of application and which involves, we are told, no essential use of normative or evaluative connotations. Rationality is, in this sense, a "broadly dispositional trait", involving a complex bundle of dispositions to behave in characteristic ways in situations the full characterization of which includes specification of further dispositions of the agent, namely, his objectives and beliefs.[6] In essence, Hempel's construal of rational explanation is, by his own admission, akin to Gilbert Ryle's notion of explanation by reference to dispositions,[7] whereby the explanandum-phenomenon is explained by being represented as a manifestation of the general disposition (of some entity involved in the explanandum-event) to act in certain characteristic ways if acted upon in certain specified ways in specified circumstances.

The "broad dispositionality" of the trait of rationality is that feature of this trait in virtue of which a variety of responses can manifest the disposition,

depending upon the circumstances. There is no single kind of action, specifiable in some more determinate fashion than simply as "rational" action, which constitutes a manifestation of rationality in any and all circumstances where such manifestation occurs. The "empirical content" of this concept would therefore have to be specified by a list of what Hempel calls "symptom sentences", each specifying either a necessary or a sufficient condition for application of the term "rationality" based upon the response of an individual in particular test conditions.[8] A Hempelian rational explanation can, thus, be said to explain a human action by invoking a symptom statement (H2) according to which the action is identified as the manifestation of the disposition of rationality in the specified circumstances, by virtue of being either a necessary or a sufficient condition for attributing the disposition of rationality to the agent in question, and a further statement (H1) which indeed attributes this disposition to the agent.

According to Ryle's conception, a wide variety of mental traits are to be understood as broadly dispositional traits in this sense. Such traits are, in Ryle's terms, highly generic or determinable, as opposed to being highly specific and determinate. They comprise abilities, tendencies, and pronenesses to do a variety of such different things that the terms which are used to refer to these traits have no corresponding episodic use.[9]

Rationality is clearly, then, not the only mentalistic trait that can be construed dispositionally, and Hempel is well aware of this point. However he expresses his awareness in a curious manner; he claims that

> An analogous diagnosis applies, incidentally, also to explanations which attribute an agent's behaviour in a given situation not to rationality and more or less explicit deliberation on his part, but to other dispositional features, such as his character and emotional make-up.[10]

Hempel suggests the example of an agent manifesting the broadly dispositional trait of fearlessness. We might suppose that, for Hempel, an agent's behaviour could be explained in terms of its fearlessness by an explanatory argument the form of which constituted only a slight modification of Hempel's schema for rational explanation, thus:

A was in situation of type C.
(H1') A was a fearless agent.
(H2') In a situation of type C, any fearless agent will do X.
Therefore A did X.

Analogously we can envisage the form of a diligent explanation, a vain explanation, and so on.

The curiosity of this view can be brought out in the following way. Broadly dispositional traits such as fearlessness are such that we can often conceive of other broadly dispositional traits which are their "opposites",

for example, cowardice. Such contrasting broadly dispositional traits can be represented equally by lists of "symptom" sentences which specify either necessary or sufficient conditions for application of the relevant disposition term based upon the response of an individual in particular test conditions. Hempel's account of rationality would seem to allow as well for the possibility of construing irrationality as a broadly dispositional trait which could conceivably characterize human beings. Certain human actions might then be correctly explained in a manner which would be schematized along Hempelian lines

A was in a situation of type C.
A was an irrational agent.
In a situation of type C, any irrational agent will do X.
Therefore A did X.

But there is an intuitive absurdity in the supposition that this sort of explanation can ever amount to an adequate account of behaviour. The supposition is absurd, not because the facts of human behaviour never conform to it in any degree--indeed, behaviour quite often seems prima facie irrational, and this may or may not be due to ignorance on the part of the observer. Rather the supposition is absurd because behaviour simply cannot be sufficiently and satisfactorily explained, or understood, in terms of irrationality. I must emphasize that neither Hempel nor anyone else of whom I am aware has ever seriously

suggested that human behaviour can be thus explained. My point is that it is implicit in Hempel's account of the form of our (occasional) understanding of human behaviour in terms of its rationality, that <u>if</u> human behaviour indeed <u>cannot</u> be adequately understood in terms of irrationality, this is merely because of what we, in our experimentation, shall <u>find</u> to be the case in the world as a matter of <u>experimental</u> fact, and not because of any conceptual absurdity in the form of "irrational" explanation schematized above. But the above form of "irrational" explanation is obviously absurd. And our feelings about its absurdity surely do not arise from the extensive accumulation of evidence currently made available by scientific experimentation, the only grounds upon which, in accord with Hempel's views, we might base such a view. Might it not perhaps rest on more fundamental metaphysical grounds than the Hempelian account can allow?

If irrationality were possible as a broadly dispositional trait of human beings, then we should be able to imagine, at least at a general and abstract level, what sort of symptom sentences might express the empirical content of the term "irrational". That is, we should be able to state what we would accept as necessary and as sufficient conditions of the manifestation of such a trait. But the most that it seems possible to say, in general, about the proposed "disposition" of irrationality is that

it would apparently manifest itself in behaviour which, in the context of its occurrence, was, at best inappropriate and irrelevant, at worst positively detrimental, to the whole network of goals, desires, interests, and even moral principles of the performing agent. Our uneasiness about the possibility that irrationality might be a dispositional trait in terms of which human behaviour could be adequately explained surely reflects the fact that behaviour is not in any way accounted for by citing either its irrelevance or its positive obstructiveness with regard to the goals and interests of the performing agent. Such behaviour has not been thereby located within any pattern of phenomena to which we might justifiably attribute the potentiality of recurrence.

The lack of explanatory power of the concept of irrationality as a possible broadly dispositional trait suggests that the concept of rationality is one which does not allow the possibility of any significant and fundamental behavioural counterinstance. It appears that the concept of rationality does not simply delineate one kind of understandable (i.e., explainable) human behaviour, as does, for example, the concept of fearlessness. I suggest that it operates instead at a much more fundamental level in our very grasp of what it is that makes any human behaviour, which is understandable as such, understandable at all.

In everyday contexts, there is a (limited) use for the concept of irrationality in regard to behaviour, namely, in regard to that behaviour which is not grasped by the observer as being appropriate in the situation of its occurrence nor for which any reasonable justification on the part of the agent can be discovered or hypothesized. But because, in everyday circumstances, the avowal of reasons may be withheld, goals may be intangible or may only be achieved in the long run, and satisfactions may be enjoyed in silence, an external observer is not in a position to ascertain with finality that for some particular action, there was in fact no reason at all--no goal, interest, desire, principle or motivating factor of any kind--which guided the performance of that action. The limitations of ordinary living are such that where reasons are actually sought for the performance of actions, only sometimes are we satisfied that they have indeed been found.

The ordinary distinction between rational behaviour and irrational behaviour accommodates itself to these limitations, and so its actual usage may be said to reveal no less about our success or failure at attempting to understand behaviour than it does about the metaphysical nature of human behaviour itself. Because of such accommodation, however, the ordinary distinction is itself limited for philosophical purposes; in particular, it is limited to an application to phenomena which must already

be classed as being human actions (in contrast to, say, mere bodily movements) on the basis of other considerations. For example, we probably rely heavily upon observations of gross bodily movements along with extensional notions of which sorts of bodily movements are likely to be genuine actions, which sorts likely to be mere physiological responses to external or internal conditions, and which may in all likelihood belong to either such group. The bodily movements which comprise human actions may have no genuine metaphysical importance in determining which phenomena are indeed human actions, even though, as a guide for everyday purposes, bodily movements doubtless serve well enough.

But even superficial everyday notions may point beyond themselves to important metaphysical considerations. Our everyday conception of rational behaviour is one which we are able to superimpose upon our conceptions of actions which are also fearless, or also cowardly, and so on. For it seems that any activity of human organisms which is understandable as a distinctively human action, though it manifest any (other) dispositional trait whatever, is understandable precisely because it is appropriate in the context of its occurrence given some desire, want, goal, interest, or principle of the performing agent. Standards of rationality, unlike, for example, standards of fearlessness, are pertinent to the understanding of all human behaviour. The metaphysical consideration to which the

everyday concept of rationality points is the notion that the only sort of activity of human organisms which is understandable as being genuinely and distinctively human behaviour is activity which is performed for the sake of desires, wants, goals, interests, and principles which are themselves (*prima facie* mentalistic) features of the behaving human organism.

Contrary to Hempel, then, I do not construe rationality as a dispositional trait, broadly so or otherwise, which merely may, or may not, characterize human beings and may, or may not, manifest itself in particular human actions. I construe rationality, metaphysically speaking, as a characteristic of such activity of human organisms as constitutes genuinely and distinctively human action. It is the characteristic of goal-directedness, or purposiveness, in the widest sense. Rationality in this sense pertains to the form of those events which are in fact human actions, namely, their *appropriateness* in the situation in which they are performed, given the wants and ends of the performing agent, and to the form of the occurrence of such events, namely, their performance *for the sake of* those ends.

At the start of the present chapter I claimed that behaviour which is to be understood in terms of its rationality can only be explained by arguments according to which the conclusions concerning the actual *performance* of

actions is grounded upon a recognition of the rationality of those actions, as expressible by laws concerning the appropriate things to be done in specified circumstances. I have also suggested that this grounding is supported by a standing presumption that human beings inherently act rationally, that is, that rational human action is a natural course of events. This is to say that the fact that an action is appropriate, in the context of its occurrence, given certain wanted ends, is no more a merely contingent regularity than is the fact that an action which is appropriate in some context actually occur in that context. This standing presumption of rationality operates to define an entire conceptual framework within which the occurrence of behaviour is to be understood in terms of the rationality of that behaviour, and an entire range of activity on the part of human organisms is thereby delineated as being metaphysically distinct from all other human activity. Rational behaviour, in this sense, is not to be contrasted with, for example, fearless behaviour, or cowardly behaviour, but with, for example, mere bodily movement and mere physiological response.

It might be thought that the Hempelian model could be modified so as to incorporate the more fundamental metaphysical construal of rationality which I am proposing while yet retaining a commitment to causal determinism. For example, suppose that Hempel's proposed premise (H1)

were revised so as to state that the agent in question was acting rationally at the time of the action to be explained, and thus, to omit any mention of supposed broadly dispositional traits. However such a revised version of premise (H1) would be redundant as an explicit premise in the explanation. In accord with the presupposition that human behaviour is inherently rational, the very giving of a "rational" explanation of an activity on the part of a human organism, as distinct from an alternative sort of explanation, itself implicitly incorporates the assessment that the activity of the human organism in question constituted a genuinely rational action, i.e., that the agent was acting rationally at the time of the action to be explained. Of course we may be dealing with an agent possessed of no mean powers of concentration who is capable of performing several actions at once. But then the revised version of (H1) would have to be further revised so as to stipulate that the agent was not only acting rationally at the time of, but also in respect of, the action to be explained, and this premise is surely as redundant, for the reasons cited above, as the initial revision.

My argument should be even more firmly secured by consideration of an analogous point which may be made with regard to explanations in terms of causally efficacious antecedent conditions. Suppose that an argument of the following form is offered in explanation of an event E:

A occurred.
The occurrence of A will cause the occurrence of E.
Therefore E occurred.

One need not include, in the explanans, an additional statement to the effect that A was causally efficacious at the time of (and in respect to) the occurrence of E. Indeed the explainer must believe that this is true in order that he be justified in offering an explanation of this form. Various exceptional circumstances, interfering factors, might frequently prevent some antecedent A from causing some event E. But the very offering (and acceptance) of such a form of explanation at all presupposes that he who offers (or accepts) it has found no positive evidence, in the case at hand, for the existence of exceptional circumstances the presence of which would have indicated that not A but some other factor among the antecedent circumstances was the cause of E. Similarly, that a rational explanation is accepted at all constitutes a (presumably reasoned) judgment on the part of one who so accepts it, that the agent was acting rationally at the time of (and in respect of) the action to be explained. The acceptance of an explanation of either type is itself an event the tacit implication of which is that there is no positive evidence indicating the irrelevance of the items cited in the explanans to the occurrence of the event to be explained.

If premise (H1) is eliminated entirely from Hempel's schematization of the form of rational explanation,

the following structure remains:

A was in a situation of type $C_1 \ldots C_n$.
(H2) In a situation of type $C_1 \ldots C_n$,
a rational agent will do X.
Therefore A did X.

Premise (H2) is highly misleading. It is surely not a merely empirical generalization about the behaviour, in a given type of situation, of a selected group of people whom we have independently ascertained to be possessed of a particular trait (viz., rationality), and whose further behaviour in the type of situation at hand need then merely be observed and recorded as the paradigm or standard of what any "rational" agent will do in that situation. Firstly, if the preceding arguments are correct, then rationality is not to be conceived as any sort of ordinary trait which human beings may or may not possess. Rather, rationality is to be conceived as being an inherent feature of all human beings. On this basis, no persons are, in principle, to be regarded as being intrinsically more rational than others.

Secondly, as a matter of empirical fact, we may find that some persons do act rationally more often than others. In accord with the presupposition that all persons are inherently rational, we would have to understand this "empirical fact" as being an indication that for some persons the natural tendency to act rationally is obstructed or interfered with less often than it is in the case of

other persons. But even if we used independent evidence to select a group of people who act rationally more often than do other persons, we could still not then rely simply on what those selected persons do in some particular type of situation $C_1 \ldots C_n$ as our standard as to what a rational agent (an agent acting rationally) will do in that type of situation. For even a human being who acts rationally more often than do most human beings, may not act rationally all the time. Consequently we could not expect that all of our "rational agents" will act in an identical fashion in the situation of type $C_1 \ldots C_n$ (even if we control the test persons so that they are possessed of identical relevant wants). Surely "the rational thing to do" is not a phrase that pertains simply to what some minimum percentage of persons happen to do.

Thirdly, a variety of persons who, on independent evidence, have been ascertained to be persons who act rationally more often than do other persons, need not all confront some particular situation $C_1 \ldots C_n$ with the same wanted ends or end-state. But the action which is rational or appropriate in a particular situation cannot be determined apart from the ends or end-state which is wanted. Thus, even if all such "rational" agents in fact continued to act rationally in the particular situation $C_1 \ldots C_n$, there need not result any uniformity in their various behaviours (aside from the rationality itself of those

behaviours) which we could then designate as being that which "a rational agent will do" in the given situation.

It seems that for the purposes of explaining behaviour in terms of its rationality, the central point which must be ascertained is the rationality of the action to be explained, and not the rationality of the agent who performed that action. As a matter of contextual presupposition, the agent is already construed as being an inherently rational agent generally (assuming the preceding arguments to have been correct), unless specific and positive evidence is regarded as proving that he acted otherwise in the particular case at hand. On the basis of such a presupposition, it is no longer a live question as to how we are to regard the agent. Rather, the question that most immediately faces us is that of how we are to regard the action which has been performed--whether as a natural manifestation of the inherent rationality of the agent, and hence, as an action which is to be explained in terms of its rationality, or as an unnatural manifestation of the obstructed rationality of the agent, and hence, as an event which is to be explained in some other terms instead, in particular, as a causal consequence of antecedent conditions. In order to answer such a question, it is necessary firstly and fundamentally to assess the rationality of the action itself in the totality of the situation at hand. And this is an assessment which, for any particular explainer, can

only justifiably be made on the basis of his own attempt to appreciate intellectually the appropriateness of the action in the total context of its occurrence.

On the basis of these considerations it would seem plausible to regard premise (H2) of the Hempelian form of rational explanation as implicitly combining two distinct considerations, one of which may be expressed as the following premise:

(H2a) In a situation of type $C_1 \ldots C_n$, the rational (or appropriate) action is X.

and the other of which may be expressed as either:

(H2b) A (generally) rational agent is one who will always (or with high probability) do X in a situation of type $C_1 \ldots C_n$.

or:

(H2c) An agent who is rational at the time of (and in regard to) a situation of type $C_1 \ldots C_n$ will do X.

It has already been argued above that a premise of the form of (H2c) would be redundant in a rational explanation by virtue of the fact that it simply reiterates something which constitutes an inherent and necessary claim implicit in the very acceptance of a rational form of explanation with regard to some particular action. Thus, if my arguments have been correct, this premise may be disregarded.

Premise (H2b) seems to be plainly irrelevant to the explanation of human behaviour in terms of its rationality, as this is herein understood. It makes no difference to the question of the rationality, and hence to the explanation, of any particular action X as performed by a particular agent that any other agent who (only) generally acts rationally will (even) always or (worse yet) with (only) high probability act in this way in a situation of the specified sort. For even after this has been said, it must still be determined whether such an agent, who generally acts rationally, would still be acting rationally if he were to do X in situations of type $C_1 \ldots C_n$. That is, premise (H2b) contributes nothing which is required for the explanation of an action in terms of its rationality that is not already contributed by component premise (H2a) which has also been extracted from Hempel's original premise (H2).

Premise (H2a) is indeed essential for a rational explanation. It is precisely this form of generalization which I have been calling the first-person practical core of "principles of action" and which, when translated into the terms of a third-person observational standpoint, constitutes the principal form of generalization that I have been taking to underlie rational explanation.

A causal generalization to the effect that people act in certain ways in situations of a particular type will

not, in and of itself, support a conclusion to the effect that that action is the appropriate thing for some particular agent to do in that situation. How is it, then, that the appropriateness of an action in a particular situation is to support the conclusion that some agent, who possessed the relevant wants, actually acted in that fashion? The possible success of this sort of reasoning requires a radically different context from that of an exhaustively causal world view. I shall now elaborate what is meant by a "standing presumption of rationality".

The presumption of the inherent rationality of action has both a metaphysical and an epistemological side. Metaphysically speaking, it is a presumption that rational action is the kind of action which is natural for human beings unless they are impeded by obstacles or interfering circumstances, just as uniform rectilinear motion is natural for moving Newtonian objects. Another way to put this point is that it is in the nature of human beings to act rationally, that the potential for rational activity is an essential defining characteristic of human beings as such. Epistemologically speaking, it is a presumption that once the agent's point of view, in particular his wants and beliefs, is reconstructed with regard to a specific situation which makes the bodily movements and/or bodily position of that agent intelligible as being the performance of a rational action in the situation, then so long as no

positive evidence indicates that the agent in question did not act in virtue of his own appreciation of the appropriateness of his own bodily condition in the situation as constituting a rational action, the occurrence of the phenomenon under consideration, now accepted as being a human action in the fullest sense, is taken to be thereby fully and adequately explained as such.

My own position is similar to, and ultimately derives from, a position defended by Charles Taylor in The Explanation of Behaviour,[11] although I diverge from Taylor's views in a number of key respects. His arguments merit special discussion in the attempt to clarify my own position. For present purposes, I shall construe my concept of rational explanation to be equivalent to Taylor's concept of what he calls "purposive explanation". Taylor construes purposive explanation as being one variety of teleological explanation, the more encompassing form. Teleological explanation is explicated as being explanation according to which events are held to occur "for the sake of" what results from them. Teleological explanations utilize teleological laws, the form of which, for Taylor, is such that the type of behaviour to be explained is construed as being dependent on the state of a system and its environment, and what the combination of the two make necessary if the end concerned is to be realized.[12] Purposive explanations of human behaviour would thus

utilize purposive laws relevant to human behaviour, according to which human behaviour would be construed as being dependent on the state of a human being and his environment, and what the combination of the two make necessary if the purpose concerned is to be achieved.

Although Taylor and I thus agree on the law-governed nature of explanations of human behaviour, his discussion of the structure of those laws is not entirely clear on all the points which I take to be important for the understanding of human behaviour in terms of its rationality. Taylor construes the antecedent of purposive laws as being

> . . . the condition of the agent having a certain intention or purpose, whether because it gives rise to a desire, or is the object of a certain policy, so that the regularity in his behaviour is conditional on the regularity of his intentions or purposes.[13]

Taylor's formulation of purposive law is ambiguous because it is not clearly a non-causal account. As expressed above it is not unlike that of many philosophers who would explain behaviour as the causal consequence of wants and beliefs.[14] For the purposes of such a thesis, wants and beliefs are construed primarily as internal states, dispositions, or events of the behaving organism, and, therefore, as part-causes within sets of causally sufficient, exhaustively determining antecedent factors. Taylor supposes the link between behaviour and antecedent intention to be non-causal because it is non-contingent, and, hence, not explicable in

terms of more basic, and non-teleological, or causal, regularities. But if it be agreed, as I do, that causal connections are connections of regular association which, as distinct from connections of merely regular association, are, in some sense, necessary, then Taylor's distinction between teleological and causal laws, on this ground, at any rate, collapses. For no good case would remain for arguing that Taylor's expressed formulation of teleological laws was not at once also causal in form.

However Taylor is correct to notice that the notion of a non-contingent link between purpose or reason, and behaviour is indeed operative in rational explanations. It is operative in the standing presumption of rationality itself. Taylor recognizes that the appeal to such natural tendencies as that of rationality is a way of attributing certain properties to the system as a whole, and not a way of identifying supposed causal antecedents. Furthermore such an appeal, on the basis of these metaphysical attributions, introduces highly significant epistemological controls into the manner of explanation. It introduces what Taylor calls an "asymmetry" in the teleological explanation of behaviour which is analogous to that introduced by the Principle of Inertia into Newtonian mechanics. Both principles distinguish, in the context of their respective systems, a kind of event which is thereby construed as being "natural", i.e., a kind of event which does not require any

further explanation beyond the point of revealing its naturalness.[15] That a body will tend to continue at rest or in uniform rectilinear motion unless disturbed (in any of an unspecified variety of ways) requires no further explanation, this natural tendency itself functioning to "explain" all instances of such inertial behaviour and to do so, in effect, by designating them to be the natural progression of events, the thing to be expected, and hence, the sort of sequence over which puzzlement and demand for explanation ought not to arise in the first place.

Similarly, that a human agent with certain wants, who believes his circumstances to be of a certain sort, and recognizes the action which he must perform in those circumstances to achieve what he wants, will endeavour to act in that way so as to achieve it--unless interfered with (in any of an unspecified variety of ways)--requires no further explanation.

Principles of action are, thus, empirical generalizations which formulate just which actions are in fact rational for human beings having particular wants and goals. And so principles of action carry explanatory force with regard to behaviour precisely in the context of an assumption that the rationality of action, when apprehended by a human being is sufficient to ensure that that human being endeavour to perform that action. It does not do so, however, by contributing to the exhaustive causal

determination of the occurrence of the action, but rather by being apprehended by a being whose essential nature encompasses the possibility of original rational activity.

In this respect, I have gone beyond Taylor's own views, for he does not appear to be committed to a "metaphysics of agency". I have, however, found this commitment to be a necessary extension of the view that behaviour is to be explained teleologically. It will be remembered that according to Taylor, the teleological form of explanation utilizes teleological laws according to which behaviour is construed as being dependent on the state of a system and its environment, and what the combination of the two make necessary if some given end is to be realized. This view shares the same problem which was found to characterize Karl Popper's thesis[16] that the behaviour of all living systems could be understood as plastically-controlled problem-solving activity. The problem is basically that any mechanical process of cause and effect may be reinterpreted as a "goal-directed" process along Taylorian lines by simply supposing that whatever happens to be its outcome has all along been the "end" toward which the events comprising the process have been "directed". In this sense, any event which is part of a causal chain of events is that which the combination of the state of the system at the time of the event and its environment make necessary if the given "end" is to be

5 5

OF/DE

realized. Thus any event whatsoever would seem to permit of a teleological form of understanding. What grounds would there be, on this basis, for plausibly maintaining that only some sequences in the world are to be understood teleologically, for example, the behaviour of living systems, and not others?

As suggested in chapter seven (pp. 270-271), this difficulty would seem to be resolved by maintaining that genuinely purposeful or rational action must be performed by a conscious being capable of acting for the sake of a wanted end which can be achieved in the future as the result, at least in part, of the effective action of that being. Additionally, such a being must also be capable of taking account of alternative future possibilities, of comprehending the effectiveness of its own activities amid specified circumstances, and of acting in a variety of ways amid (at least some) specified sets of circumstances, both internal as well as external. The standing presumption of the inherent rationality of human beings would seem frivolous if it had only an epistemological dimension; at bottom it is a metaphysical presumption concerning at least a part of the essential nature which is attributed to any human being whose behaviour is to be understood in terms of its rationality. And this presumption is construed as being generally applicable, until positive evidence proves it to be inapplicable in particular cases.

One further difficulty with regard to Taylor's conception of teleological laws deserves mention. Contrary to his conclusions, principles of action may indeed be subsumable under yet more basic empirical laws. However the empirical laws under which they may be subsumed need not be causal in form. Indeed they cannot be causal in form for then subsumption is not logically possible. Subsumption of such laws might occur in the following sort of way: if a particular type of action, A_1, were an instance of a more general class of actions, A_2, and if a particular type of situation, S_1, were an instance of a more general class of situations, S_2, then an empirical law expressing the appropriateness of actions of type A_1 in situations of type S_1 might be subsumable under a more encompassing empirical law expressing the appropriateness of actions of type A_2 in situations of type S_2. In this respect, then, rational laws, or principles of action, do not, as Taylor claims of teleological laws in general, represent "a deviation from the modern norm and a throw-back to an earlier type of explanation".[17] They leave open, as much as do causal laws, the possibility that a more basic set of laws--<u>of the same form</u>--will be found.

The asymmetry which Taylor attributes to the purposive form of explanation, in virtue of its being an instance of the more general teleological form of explanation, is more aptly characterized as an asymmetry

in the application of the notion of purposiveness, or rationality, to phenomena. Any theory of human behaviour which constitutes an attempt to explain and understand human behaviour as a distinctively rational process, must incorporate as a basic principle the notion of rationality as the natural tendency, the expected direction of human behaviour, only deviations from which require further explanation in terms of antecedent conditions thereby construed as disturbing or interfering factors. This basic principle, like the Principle of Inertia within the context of Newtonian Mechanics, is never open to direct empirical test or refutation within the context of the form of understanding of human behaviour which I am delineating. One might wish to say that it is therefore neither true nor false, but that, considered epistemologically, it has instead the status of a "rule". As a rule, the standing presumption of human rationality enjoins us to deem adequate and satisfactory an explanation of human action (or, more precisely, human endeavour) which cites as explanatory factors such as the agent's beliefs and wants as together would render the agent's intended action appropriate, or rational, in the circumstances that the agent envisages to obtain.

Notice that for an adequate explanation of human behaviour in terms of its rationality, the agent's action must be recognized by the explainer-observer to be rational

FOOTNOTES

[1]Laws and Explanation in History (Oxford, 1957), esp. ch. 5; cf. also his "The Historical Explanation of Actions Reconsidered", in Sidney Hook, ed., Philosophy and History: A Symposium (New York, 1963), pp. 105-135.

[2]Cf. Carl Hempel, "Aspects of Scientific Explanation", §10: The Concept of Rationality and the Logic of Explanation by Reasons, Aspects of Scientific Explanation (New York, 1965), esp. p. 470; also Alan Donagan, "The Popper-Hempel Theory Reconsidered", in William Dray, ed., Philosophical Analysis and History (New York, 1966), pp. 127-159, esp. pp. 155-157.

[3]Hempel, loc. cit.; Donagan, op. cit., p. 155.

[4]Cf. James Leach, "Dray on Rational Explanation", Philosophy of Science, vol. 33 (1966), p. 65; also Hempel, op. cit., p. 471.

[5]Loc. cit.

[6]Ibid., pp. 472-473.

[7]The Concept of Mind (London, 1949).

[8]Op. cit., pp. 460-461.

[9]Ryle, op. cit., pp. 42-45; 117-119.

[10]"Explanation in Science and in History", in William Dray, ed., Philosophical Analysis and History, p. 118. Italics my own.

collected. If the given behaviour is to be explained in terms of its rationality then the explainer must seek further evidence regarding the agent and his circumstances in order to determine if the behaviour was rational after all.

Of course, for an explainer to appreciate the appropriateness of someone else's behaviour, he need only take that person's point of view as "given" in the course of his reasoning; he need not, as some mistakenly suppose,[19] commit himself to the agent's very beliefs and goals, i.e., he need not actually believe or want as the agent himself actually believed and wanted. Only to the connection between this "given", in the specified circumstances, and the actually ensuing action must there be an attempt at anything like "commitment". For it is this connection in which consists the rationality or lack of it that the action had in the context of its occurrence. If there is any philosophically valuable point to be made by "empathy" theorists in regard to rational explanation, it would seem to consist in recognition of this active intellectual "appreciation" which an explainer must have of a given action in order that he may justifiably deem himself to have satisfactorily explained, and therefore, understood, that action in terms of its rationality.

The intellectual connection between action and the total situation in which it occurs, when it is one of

appropriateness, makes a claim upon the assent of all of us in just the same fashion as does any logical implication by the premises of an argument. When one fully understands a particular set of premises and fully understands a further statement which is said to be a logical deduction from those premises, yet one does not grasp this further statement as being logically entailed by the given premises, then one is not warranted in accepting any such purported argument as being in fact logically adequate. One cannot justifiably believe in logical implications on grounds of hearsay alone. A similar point holds with regard to the understanding of human behaviour in terms of its rationality, and to the belief that particular actions to be explained were in fact rational in the envisaged contexts of their performance. The explainer's own intellectual capacity for assessing action is ultimately his best standard as to the adequacy of a given set of factors adduced to reveal the rationality, and, hence, explain the performance, of a particular instance of a human behaviour. For the explainer is, himself, also a rational agent.

It is thus a standing presumption that rational activity is an inherent and essential characteristic of human beings, and therefore that human behaviour is naturally rational, which legitimizes logical moves from principles of action, formulating the <u>appropriateness</u>, or <u>rationality</u> of various types of actions in given situations,

to laws expressing the actual performance of those types of actions in those situations. Because rationality is herein understood as a defining characteristic of human behaviour as such, it cannot be taken to constitute a mere broadly dispositional trait of perhaps this or that human being whose behaviour is then explained in terms of this dispositional trait in such cases as it may be supposed as having been operative.

One final comment about the rational model for interpreting human behaviour is in order. Following Körner, I have suggested that no theory in itself entails the limits of its applicability. It is only in applying, or refusing to apply, a theory to given cases that we, in effect, presuppose what these limits are to be. The rational model for interpreting human behaviour has been presented in the present context as if there were no limitations to its applicability to human activity with the exception of physiological activity and mere bodily response, as in the case of the knee-jerk reflex. However psychiatry appears to present us with explanations of human actions in terms of, for example, phobias and neuroses, which would not appear to conform to the rational model. In such cases the behaviour is either not explained as the endeavour toward action intended by the agent, or it is not explained by appeal to anything that constitutes a belief or want of the agent.

At any rate, this is what at first sight appears to be the case with regard to many psychiatric explanations. Let us call such psychiatric explanations of behaviour which do not seem to represent the behaviour as being rational, as "abnormal" psychiatric explanations, for they pertain to a branch of psychiatry that deals with apparently abnormal behaviour. Presumably such explanations have gained credence empirically through satisfactory applications to particular cases.

There are, at least, two ways of acknowledging the possibility of such apparently non-rational explanations with regard to human behaviour without abandoning anything that has been maintained here in connection with the understanding of human behaviour in terms of its rationality. On the one hand, we can acknowledge the applicability of a duality of frameworks for understanding various human actions. On the other hand we can defend the unlimited applicability of the rational model by postulating that all apparently irrational or non-rational behaviour, at least at the level of gross overall characteristics,[20] is only apparently so, the mistaken appearance due to ignorance on our parts of the real underlying <u>reasons</u> on the parts of the agents for such behaviour. It is a matter of debate as to which of these two views is a proper interpretation of Freud's work regarding explanation of behaviour in terms of unconscious conditions.[21] In any case, Dray, for one, is

quite content with an explanatory duality at the level of genuine human behaviour, although for Dray the two relevant models--the rational and the causal--constitute alternatives to be sought in a certain order. His way of putting the point is that "We give reasons if we can, and turn to empirical laws if we must",[22] a choice of terms which reflects Dray's view that only explanation by causes, and not that by reasons, invokes "empirical laws", a view which I have rejected.

Empirical evidence may appear to favour such a dualistic conception of human behaviour, but the concept of a unitary approach to human behaviour has a compelling intellectual attractiveness. Such a view is not implausible even as an account of abnormal psychiatric explanations. For example, as G. J. Warnock notes, explanations of actions which undermine and replace the explanations based on reasons which people give for their own behaviour are at least sometimes still explanations in terms of reasons. The only difference is that these are now reasons which the agent simply does not acknowledge to others, or even to himself, perhaps because they are of a less reputable kind. Thus they are still reasons of the agent, but reasons of which the agent may not be wholly or even partly conscious.[23] Since the explanation of human behaviour in terms of its rationality does not require that an agent have <u>consciously</u> apprehended the rationality of some

possible action for him in order to have acted on the basis of such apprehension, then this interpretation of abnormal psychiatric explanation makes it quite amenable to reconstruction in accord with the rational form of explanation presented here.

Suppose that behaviour seems to be irrational or non-rational because the reasons given by the agent in question are inadequate as reasons. This might mean one of two things: either they do not logically imply the appropriateness of the behaviour (all relevant types of conditions being assumed), or they do logically imply this yet are so implausible, pernicious, or outrageous as reasons that we cannot imagine how the agent could have been committed to them. In the latter sort of case, the full explanation of behaviour, in order to finally satisfy us, will have supplemented the above information with, as J. F. Thomson notes, further information regarding such factors as that of early upbringing which explains how the agent came to be influenced by the unusual reason. And these are factors which are different in kind from factors that the agent himself could mention as "reasons" for his action.[24] However such a supplemented account does not negate the appropriateness of the agent's behaviour given his unusual wants or beliefs; it merely adds an explanation of why those wants or beliefs characterized the agent in the first place. For example a child may assiduously avoid

stepping on cracks in the sidewalk because he sincerely believes that should he do so, harm will befall him. He believes this because he has been told this by his trusted friends. Such action is not irrational in terms of what the child believes and wants; it is merely ill-founded.

What if the reasons given by the agent for his action simply do not logically imply the appropriateness of the behaviour, even when all relevant types of conditions have been assumed. (That is, if the agent cites a want as his "reason" for acting in a certain way, this explanation must be supplemented by an assumption regarding circumstances which the agent believes to obtain and cause-effect relationships pertinent thereto.) It seems then, that we shall have to search for further explanatory factors of some sort, whether of the rational or causal variety. For, as was claimed earlier in this chapter, behaviour is surely not in any way accounted for by citing "explanatory" factors in terms of which it appears either irrelevant or positively obstructive with regard to the wants of the performing agent. Thus this sort of example appears to devolve into one of the three sorts of cases noted above: either it is a case in which a rational explanation can be found, but with the unusual feature that either the reasons are at variance with the conscious "reasons" of the agent or the reasons themselves are so puzzling that the entire explanation does not seem satisfactory until those reasons

are explained, or it is a case in which no rational explanation can be found but a causal explanation seems satisfactory.

In no case does there seem to be any serious threat to the adequacy of the explanation of human behaviour in terms of its rationality, as schematized and explicated in this dissertation, for at least a generous portion, and perhaps all of what is called human behaviour in any sense implying intentionality, quasi-intentionality, reason, purpose, goal, want, interest, desire, motive, or the like.

FOOTNOTES

[1]Laws and Explanation in History (Oxford, 1957), esp. ch. 5; cf. also his "The Historical Explanation of Actions Reconsidered", in Sidney Hook, ed., Philosophy and History: A Symposium (New York, 1963), pp. 105-135.

[2]Cf. Carl Hempel, "Aspects of Scientific Explanation", §10: The Concept of Rationality and the Logic of Explanation by Reasons, Aspects of Scientific Explanation (New York, 1965), esp. p. 470; also Alan Donagan, "The Popper-Hempel Theory Reconsidered", in William Dray, ed., Philosophical Analysis and History (New York, 1966), pp. 127-159, esp. pp. 155-157.

[3]Hempel, loc. cit.; Donagan, op. cit., p. 155.

[4]Cf. James Leach, "Dray on Rational Explanation", Philosophy of Science, vol. 33 (1966), p. 65; also Hempel, op. cit., p. 471.

[5]Loc. cit.

[6]Ibid., pp. 472-473.

[7]The Concept of Mind (London, 1949).

[8]Op. cit., pp. 460-461.

[9]Ryle, op. cit., pp. 42-45; 117-119.

[10]"Explanation in Science and in History", in William Dray, ed., Philosophical Analysis and History, p. 118. Italics my own.

[11](London, 1964). Dray also defends a standing presumption that "people act for sufficient reason". Cf. Laws and Explanation in History, pp. 137-139.

[12]Ibid., pp. 5-6, 9.

[13]Ibid., p. 36. Italics my own.

[14]For a representative list of such philosophers, see chapter five, footnote #1.

[15]Taylor, op. cit., pp. 17-18, 20-21, 23-24.

[16]See discussion, chapter seven, pp. 283-284.

[17]Op. cit., p. 21.

[18]Foundations of Historical Knowledge (New York, 1965), pp. 190-194.

[19]Cf. Ibid., p. 193.

[20]I have already suggested that behaviour, like physical phenomena, may not be explainable deterministically down to the very last minute detail, that there may be a rationally undetermined "fringe" of respects concerning any action which cannot be appreciated as having a regular relationship of appropriateness to anything else pertaining to the agent or his behaviour. But this is not to suggest that a behavioural event could be rationally undetermined in all respects, particularly in its gross overall character, and still be regarded as a genuine human action (rather than, say, as a bodily convulsion or seizure).

[21]Cf. R. S. Peters, The Concept of Motivation (London, 1958), pp. 62-94, esp. pp. 62-71. Peters argues that Freud insisted on a radical difference between explanation in terms of conscious processes and that in terms of unconscious processes. For a contrary viewpoint, cf. Jerome S. Bruner, "The Freudian Conception of Man", in May Brodbeck, ed.,

Readings in the Philosophy of the Social Sciences (New York, 1968), pp. 705-711. Bruner seems to suggest that, among other collapsed distinctions, Freud took the apparently irrational in behaviour to be continuous with (and understood in the same terms as) the apparently rational. For example, Bruner says, "If today people are reluctant to report their dreams with the innocence once attached to such recitals, it is again because Freud brought into common question the discontinuity between the rational purposefulness of waking life and the seemingly irrational purposelessness of fantasy and dream". (p. 709)

[22]Laws and Explanation in History, pp. 137-139. Quote is from p. 138.

[23]See discussion between G. J. Warnock, P. F. Strawson, and J. F. Thomson, "Determinism", in D. F. Pears, ed., Freedom and the Will (London, 1965), p. 62.

[24]Ibid., p. 58.

BIBLIOGRAPHY

Ackermann, Robert. Philosophy of Science. New York: Pegasus, 1970.

Alexander, Peter. "Rational Behaviour and Psychoanalytic Explanation". In: Norman S. Care and Charles Landesman, eds. Readings in the Theory of Action. Bloomington: Indiana University Press, 1968; pp. 159-178. Reprinted from: Mind, LXXI (1962).

Alston, William P. "Wants, Actions, and Causal Explanation". In: Hector-Neri Castaneda, ed. Intentionality, Minds, and Perception. Detroit: Wayne State University Press, 1967; pp. 301-341. Comments by Keith Lehrer, pp. 342-350, and rejoinder by William P. Alston, pp. 351-356.

Anscombe, G. E. M. Intention. Oxford: Basil Blackwell, 1957 (second edition, 1963).

Aune, Bruce. "Hypotheticals and 'Can': Another Look", Analysis, vol. 27, no. 6 (June 1967), pp. 191-195.

Austin, J. L. "Ifs and Cans". Philosophical Papers. Oxford: Clarendon Press, 1961; pp. 153-180. Reprinted from: Proceedings of the British Academy, 1956.

Baier, Kurt. The Moral Point of View. New York: Random House, 1958.

_____. "Responsibility and Action". In: Myles Brand, ed. The Nature of Human Action. Glenview, Ill.: Scott, Foresman and Company, 1970, pp. 100-116.

Beck, Clive. "Utterances Which Incorporate a Value Statement", American Philosophical Quarterly, vol. 4, no. 4 (October 1967), pp. 291-299.

Bennett, Jonathan. Rationality. New York: Humanities Press, 1964.

Bergmann, Gustav. "Imperfect Knowledge". In: May Brodbeck, ed. Readings in the Philosophy of the Social Sciences. New York: The Macmillan Company, 1968; pp. 415-436. Reprinted from: Gustav Bergmann, Philosophy of Science. Madison: University of Wisconsin Press, 1957.

_____. "Purpose, Function, Scientific Explanation". In: May Brodbeck, ed. Readings in the Philosophy of the Social Sciences. New York: The Macmillan Company, 1968; pp. 211-223.

Bergson, Henri. Time and Free Will. F. L. Pogson, trans. London: George Allen & Unwin, 1910.

Berofsky, Bernard. Determinism. Princeton University Press. 1971.

_____. "Purposive Action". American Philosophical Quarterly, vol. 7, no. 4 (October 1970), pp. 311-320.

Binkley, Robert. "A Theory of Practical Reason". Philosophical Review, vol. 74 (1965), pp. 423-448.

Black, Max. "The Gap Between 'Is' and 'Should'". In: W. D. Hudson, ed. The Is-Ought Question. London: Macmillan, 1969; pp. 99-113. Reprinted from: The Philosophical Review, LXXIII (1964), pp. 165-181.

_____. "Making Something Happen". In: Sidney Hook, ed. Determinism and Freedom in the Age of Modern Science. New York: New York University Press, 1958; pp. 15-30.

Blanshard, Brand. "The Case for Determinism". In: Sidney Hook, ed. Determinism and Freedom in the Age of Modern Science. New York: New York University Press, 1958; pp. 3-15.

Bohm, David. Causality and Chance in Modern Physics. Philadelphia: University of Pennsylvania Press, 1957.

Boyle, Joseph M., Jr., Germain Grisez, Olaf Tollefsen, "Determinism, Freedom, and Self-Referential Arguments", Review of Metaphysics, vol. XXVI, no. 1 (September 1972), pp. 3-37.

Bradley, M. C. "A Note on Mr. MacIntyre's Determinism". In: Bernard Berofsky, ed. Free Will and Determinism. New York: Harper & Row, 1966; pp. 256-264. Reprinted from: Mind, LXVIII, no. 272 (October 1959), pp. 521-525.

Braithwaite, Richard Bevan. Scientific Explanation. New York: Harper & Row, 1960. (First published New York: Cambridge University Press, 1953.)

Brand, Myles. "Causes of Actions". Journal of Philosophy, vol. LXVII, no. 31 (November 5, 1970), pp. 932-947.

Brandt, Richard. "Determinism and the Justifiability of Moral Blame". In: Sidney Hook, ed. Determinism and Freedom in the Age of Modern Science. New York: New York University Press, 1958; pp. 137-143.

_____ and Jaegwon Kim. "Wants as Explanations of Actions". In: Norman S. Care and Charles Landesman, eds. Readings in the Theory of Action. Bloomington: Indiana University Press, 1968; pp. 199-213. Reprinted from: The Journal of Philosophy, LX (1963), pp. 425-435.

Broad, C. D. "Determinism, Indeterminism, and Libertarianism". In: Bernard Berofsky, ed. Free Will and Determinism. New York: Harper & Row, 1966; pp. 135-159. Reprinted from: C. D. Broad. Ethics and the History of Philosophy. London: Routledge & Kegan Paul, 1952; pp. 195-217.

_____. The Mind and Its Place in Nature. London: Routledge & Kegan Paul, 1925.

Brown, Robert. Explanation in Social Science. Chicago: Aldine Publishing Company, 1963.

Bruner, Jerome S. "The Freudian Conception of Man". In: May Brodbeck, ed. Readings in the Philosophy of the Social Sciences. New York: The Macmillan Company, 1968; pp. 705-711. Reprinted from: Daedalus, vol. 87 (1958), pp. 77-84.

Campbell, C. A. "Is 'Freewill' A Pseudoproblem?" Mind, vol. LX, no. 240 (October 1951), pp. 441-465.

_____. On Selfhood and Godhood. London: George Allen & Unwin Ltd., 1957. Revised and expanded from The Gifford Lectures delivered at the University of St. Andrews, 1953-54 and 1954-55.

Castaneda, Hector-Neri. "Imperatives, Decisions, and 'Oughts': A Logicometaphysical Investigation". In: Hector-Neri Castaneda, ed. Morality and the Language of Conduct. Detroit: Wayne State University Press, 1963.

Chisholm, Roderick M. "Freedom and Action". In: Keith Lehrer, ed. Freedom and Determinism. New York: Random House, 1966; pp. 11-44.

_____. "'He Could Have Done Otherwise'". In: Myles Brand, ed. The Nature of Human Action. Glenview, Ill.: Scott, Foresman and Company, 1970, pp. 293-301. Reprinted with revisions from: The Journal of Philosophy, LXIV, no. 13 (July 6, 1967), pp. 409-417.

_____. "On the Logic of Intentional Action". In: Robert Binkley, Richard Bronaugh, and Ausonio Marras, eds. Agent, Action, and Reason. Toronto: University of Toronto Press, 1971; pp. 38-69. With comments by Bruce Aune, pp. 69-75, and Reply by Roderick Chisholm, pp. 76-80.

Churchland, Paul M. "The Logical Character of Action-Explanations". The Philosophical Review, vol. LXXIX, no. 2 (April 1970), pp. 214-236.

Danto, Arthur C. "Basic Actions". Ind: Myles Brand, ed. The Nature of Human Action. Glenview, Ill: Scott, Foresman, and Company, 1970; pp. 255-266. Reprinted from: American Philosophical Quarterly, II (April 1965), pp. 141-48.

Daveney, T. F. "Choosing". In: Myles Brand, ed. The Nature of Human Action. Glenview, Ill.: Scott, Foresman and Company, 1970, pp. 82-90. Reprinted from: Mind, LXXIII (1964), pp. 515-526.

Davidson, Donald. "Actions, Reasons, and Causes". In: May Brodbeck, ed. Readings in the Philosophy of the Social Sciences. New York: The Macmillan Company, 1968; pp. 44-58. Reprinted from: The Journal of Philosophy, LX, no. 23 (November 7, 1963), pp. 685-700.

_____. "Mental Events". In: Lawrence Foster & J. W. Swanson, eds. Experience & Theory. University of Massachusetts Press, 1970; pp. 79-101.

Donagan, Alan. "The Popper-Hempel Theory Reconsidered". In: William Dray, ed. Philosophical Analysis and History. New York: Harper & Row, 1966, pp. 127-159. Reprinted from: History and Theory, vol. IV, no. 1 (1964); pp. 3-26, under the title "Historical Explanation: The Popper-Hempel Theory Reconsidered".

Dray, William. "The Historical Explanation of Actions Reconsidered". In: Sidney Hook, ed. Philosophy and History: A Symposium. New York University Press, 1963; pp. 105-135.

_____. Laws and Explanation in History. Oxford: Clarendon Press, 1957.

Ducasse, C. J. "Determinism, Freedom, and Responsibility". In: Sidney Hook, ed. Determinism and Freedom in the Age of Modern Science. New York: New York University Press, 1958; pp. 147-157.

_____. "Explanation, Mechanism, and Teleology". In: Herbert Feigl and Wilfrid Sellars, eds. Readings in Philosophical Analysis. New York: Appleton-Century-Crofts, Inc., 1949; pp. 540-544. Reprinted from: The Journal of Philosophy, vol. 23 (1926).

Edwards, Paul. "Hard and Soft Determinism". In: Sidney Hook, ed. Determinism and Freedom in the Age of Modern Science. New York: New York University Press, 1958; pp. 105-109.

Farber, I. E. "Personality and Behavioural Science". In: May Brodbeck, ed. Readings in the Philosophy of the Social Sciences. New York: The Macmillan Company, 1968; pp. 145-179. Reprinted from: "A Framework for the Story of Personality as a Behavioural Science". In: P. Worchel and D. Byrne, eds. Personality and Personality Change. New York: John Wiley & Sons, 1964; pp. 3-37.

Feigl, Herbert. The "Mental" and the "Physical": The Essay and a Postscript. Minneapolis: University of Minnesota Press, 1967. The essay reprinted from: Herbert Feigl, Michael Scriven, and Grover Maxwell, eds. Minnesota Studies in the Philosophy of Science, vol. II. Minneapolis: University of Minnesota Press, 1958.

Flew, Antony. "On Not Deriving 'Ought' From 'Is'". In: W. D. Hudson, ed. The Is-Ought Question. London: Macmillan, 1969; pp. 135-143. Reprinted from: Analysis 25 (1964).

Foot, Philippa. "Free Will as Involving Determinism". In: Bernard Berofsky, ed. Free Will and Determinism. New York: Harper & Row, 1966, pp. 95-108. Reprinted from: The Philosophical Review, LXVI, no. 4 (October 1957), pp. 439-450.

Frank, Philipp. Philosophy of Science. Englewood Cliffs: Prentice-Hall, 1957.

Frankfurt, Harry G. "Freedom of the Will and the Concept of a Person", The Journal of Philosophy, vol. LXVIII, no. 1 (January 14, 1971), pp. 5-20.

Geach, P. T. "Ascriptivism". In: Myles Brand, ed. The Nature of Human Action. Glenview, Ill.: Scott, Foresman and Company, 1970; pp. 117-120. Reprinted, and slightly revised, from: The Philosophical Review, LXIX (1960), pp. 221-225.

Gibson, Quentin. The Logic of Social Inquiry. New York: The Humanities Press, 1960.

Ginet, Carl. "Might We Have No Choice". In: Keith Lehrer, ed. Freedom and Determinism. New York: Random House, 1966; pp. 87-104.

Goldman, Alvin I. A Theory of Human Action. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1970.

Grunbaum, Adolf. "Causality and the Science of Human Behaviour". In: Herbert Feigl and May Brodbeck, eds. Readings in the Philosophy of Science. New York: Appleton-Century-Crofts, Inc., 1953; pp. 766-778. Reprinted from: American Scientist, v. 40 (1952).

_____. "Free Will and Laws of Human Behaviour". American Philosophical Quarterly, vol. 8, no. 4 (October 1971), pp. 299-317.

Gunner, D. L. "Professor Smart's 'Sensations and Brain Processes'". In: C. F. Presley, ed. The Identity Theory of Mind. St. Lucia, Queensland: University of Queensland Press, 1967; pp. 1-20. Originally presented at First Annual Congress of the Australasian Association of Philosophy, University of Queensland (Brisbane), 1964.

Hamlyn, D. W. "Behaviour". In: V. C. Chappell, ed. The Philosophy of Mind. Englewood Cliffs: Prentice-Hall, 1962; pp. 60-73. Reprinted from: Philosophy, XXVIII (1953), pp. 132-145.

Hampshire, Stuart. Thought and Action. London: Chatto and Windus, 1960.

Hanson, Norwood Russell. Patterns of Discovery. Cambridge: University Press, 1965.

Hare, R. M. Freedom and Reason. New York: Oxford, 1965.

_____. The Language of Morals. New York: Oxford University Press, 1964 (First Published 1952).

_____. "Wanting: Some Pitfalls". In: Robert Binkley, Richard Bronaugh, and Ausonio Marras, eds. Agent, Action, and Reason. Toronto: University of Toronto Press, 1971; pp. 81-97. With comments by David Gauthier, pp. 98-108, and David Pears, pp. 108-127.

Harmon, Gilbert H. "Knowledge, Reasons, and Causes". The Journal of Philosophy, vol. LXVII, no. 21 (November 5, 1970), pp. 841-855.

Hempe, Carl. "Aspects of Scientific Explanation". In: Aspects of Scientific Explanation. New York: The Free Press, 1965; pp. 463-487.

_____. "Explanation in Science and in History". In: William Dray, ed. Philosophical Analysis and History. New York: Harper & Row, 1966; pp. 95-126. Reprinted from: R. G. Colodny, ed. Frontiers of Science and Philosophy. Pittsburgh, 1962; pp. 9-33.

_____. "The Function of General Laws in History". In: Aspects of Scientific Explanation. New York: The Free Press, 1965; pp. 231-243. Reprinted from: The Journal of Philosophy, vol. 39 (1942), pp. 35-48.

_____. "Reasons and Covering Laws in Historical Explanation". In: Sidney Hook, ed. Philosophy and History: A Symposium. New York University Press, 1963; pp. 143-163.

_____. "Rational Action". In: Norman S. Care and Charles Landesman. Readings in the Theory of Action. Bloomington: Indiana University Press, 1968; pp. 281-305. Reprinted from: Proceedings and Addresses of the American Philosophical Association, XXXV (1962).

_____. "Some Reflections on 'The Case for Determinism'". In: Sidney Hook, ed. Determinism and Freedom in the Age of Modern Science. New York: New York University Press, 1958; pp. 157-163.

_____ and Paul Oppenheim. "Studies in the Logic of Explanation". In: Aspects of Scientific Explanation. New York: The Free Press, 1965; pp. 245-290. Reprinted from: Philosophy of Science, vol. 15 (1948), pp. 135-175.

Herbst, Peter. "A Critique of the Materialist Identity Theory". In: C. F. Presley, ed. The Identity Theory of Mind. St. Lucia, Queensland: University of Queensland Press, 1967; pp. 38-64. Originally presented at First Annual Congress of the Australasian Association of Philosophy, University of Queensland (Brisbane), 1964.

Hobart, R. E. "Free Will as Involving Determination and Inconceivable Without It". In: Bernard Berofsky, ed. Free Will and Determinism. New York: Harper & Row, 1966; pp. 63-95. Reprinted from:- Mind, XLIII, no. 169 (January 1934), pp. 1-27.

Hook, Sidney. "Necessity, Indeterminism, and Sentimentalism". In: Sidney Hook, ed. Determinism and Freedom in the Age of Modern Science. New York: New York University Press, 1958; pp. 167-180.

Hospers, John. "What is Explanation?" In: Antony Flew, ed. Essays in Conceptual Analysis. London: Macmillan & Company, 1956; pp. 94-119. Revised and reprinted from: Journal of Philosophy, v. XLIII (1946).

_____. "What Means this Freedom?" In: Bernard Berofsky, ed. Free Will and Determinism. New York: Harper & Row, 1966; pp. 26-45. Reprinted from: Sidney Hook, ed. Determinism and Freedom in the Age of Modern Science. New York: Collier Books, 1961; pp. 126-142.

Johnson, W. E. Logic, Part III: The Logical Foundations of Science. New York: Dover Publications, 1921.

Kaiser, D. Nolan. "Philippa Foot and the Concepts of Law, Intention, and Accident", Mind, vol. LXXVIII, no. 310 (April, 1969), pp. 273-277.

Kane, R. H. "Minds, Causes, and Behaviour". The Review of Metaphysics, vol. XXIV, no. 2 (December 1970), pp. 302-334.

Kant, Immanuel. Foundations of the Metaphysics of Morals trans. by Lewis White Beck. Indianapolis: Bobbs-Merrill Company, (Liberal Arts), 1959. Originally published 1785.

Kenny, Anthony. Action, Emotion and Will. London: Routledge & Kegan Paul, 1963.

Kneale, William. Probability and Induction. Oxford: Clarendon Press, 1949.

Korner, Stephen. Experience and Theory. London: Routledge & Kegan Paul, 1966.

Kovesi, Julius. Moral Notions. New York: Humanities Press, 1967.

Kuhlenbeck, Hartwig. Mind and Matter. Basel (Switzerland): S. Karger, 1961.

Landesman, Charles. "The New Dualism in the Philosophy of Mind", Review of Metaphysics, 19 (1965), pp. 329-345.

Leach, James. "Dray on Rational Explanation", Philosophy of Science, vol. 33 (1966), pp. 61-69.

Lehrer, Keith. "An Empirical Disproof of Determinism?" In: Keith Lehrer, ed. Freedom and Determinism. New York: Random House, 1966; pp. 175-202.

Leibniz, Gottfried Wilhelm. Basic Writings: Discourse on Metaphysics, Correspondence with Arnauld, and Monadology. Introduction by Paul Janet; trans. by George R. Montgomery. La Salle: Open Court, 1962 (2nd Edition).

Levison, A. B. and I. Thalberg. "Essential and Causal Explanations of Action". Mind, vol. LXXVIII, no. 309 (January, 1969), pp. 91-101.

Locke, John. An Essay Concerning Human Understanding. In: Great Books of the Western World, vol. 35. Chicago: Encyclopedia Britannica, 1952; pp. 85-395, esp. Bk. II, ch. xxi, pp. 178-200.

Louch, A. R. Explanation and Human Action. Berkeley: University of California Press, 1969.

Mace, C. A. "Mechanical and Teleological Causation". In: Herbert Feigl and Wilfrid Sellars, eds. Readings in Philosophical Analysis. New York: Appleton-Century-Crofts, Inc., 1944; pp. 534-539. Excerpted and reprinted from: Aristotelian Society, Supplementary Volume 14 (1935).

MacIntyre, A. C. "Determinism". In: Bernard Berofsky, ed. Free Will and Determinism. New York: Harper & Row, 1966; pp. 240-256. Reprinted from: Mind, LXVI, no. 261 (January, 1957), pp. 28-41.

_____. "The Idea of a Social Science". The Aristotelian Society, Supplementary Volume XLI (1967), pp. 95-114.

Mackay, Donald M. Freedom of Action in a Mechanistic Universe. Cambridge: Cambridge University Press, 1967. (Twenty-first Eddington Memorial Lecture, delivered at Cambridge 17 Nov. 67).

_____. "Mindlike Behaviour in Artefacts". In: Kenneth M. Sayre and Frederick J. Crosson, eds. The Modeling of Mind. New York: Simon and Schuster, 1963; pp. 225-241. Reprinted from: The British Journal for the Philosophy of Science, v. II (1951-52), pp. 105-21.

_____. "On the Logical Indeterminacy of a Free Choice". Mind, vol. 69 (1960), pp. 31-40.

Macklin, Ruth. "Action, Causality, and Teleology". British Journal for the Philosophy of Science, vol. 19, no. 4 (February 1969), pp. 301-316.

_____. "Reasons vs. Causes in Explanation of Action". Philosophy and Phenomenological Research, vol. XXXIII, no. 1 (September 1972), pp. 78-89.

Madell, Geoffrey. "Action and Causal Explanation". Mind, vol. LXXVI, no. 301 (January, 1967), pp. 34-48.

Margolis, Joseph. "Puzzles Regarding Explanation by Reasons and Explanation by Causes". Journal of Philosophy, vol. LXVII, no. 7 (April 9, 1970), pp. 187-195.

McCall, Storrs. "Ability as a Species of Possibility". In: Myles Brand, ed. The Nature of Human Action. Glenview, Ill.: Scott, Foresman and Company, 1970; pp. 139-147.

Medlin, Brian. "Ryle and the Mechanical Hypothesis". In: C. F. Presley, ed. The Identity Theory of Mind. St. Lucia, Queensland: University of Queensland Press, 1967; pp. 94-150. Enlarged version of paper originally presented at First Annual Congress of the Australasian Association of Philosophy, University of Queensland (Brisbane), 1964.

Meehl, Paul E. "Psychological Determinism and Human Rationality: A Psychologist's Reactions to Professor Karl Popper's 'Of Clouds and Clocks'". In: Michael Radner and Stephen Winokur, eds. Minnesota Studies in the Philosophy of Science, vol. IV. Minneapolis: University of Minnesota Press, 1970; pp. 310-372.

Mehlberg, Henryk. "Physical Laws and Time's Arrow". In: Herbert Feigl and Grover Maxwell, eds. Current Issues in the Philosophy of Science. New York: Holt, Rinehart and Winston, 1961; pp. 105-138.

Melden, A. I. Free Action. London: Routledge & Kegan Paul, 1961.

Mill, John Stuart. "The Freedom of the Will". In: Bernard Berofsky, ed. Free Will and Determinism. New York: Harper & Row, 1966; pp. 159-174. Reprinted from: John Stuart Mill. An Examination of Sir William Hamilton's Philosophy. New York: Holt, Rinehart and Winston, 1874; vol. II, pp. 272-290.

Mischel, Theodore. "Psychology and Explanations of Human Behaviour". In: Norman S. Care and Charles Landesman. Readings in the Theory of Action. Bloomington: Indiana University Press, 1968; pp. 214-237. Reprinted from: Philosophy and Phenomenological Research, XXIII (1963).

Munitz, Milton K. "The Relativity of Determinism". In: Sidney Hook, ed. Determinism and Freedom in the Age of Modern Science. New York: New York University Press, 1958; pp. 63-69.

Nagel, E. "Mechanistic Explanation and Organismic Biology". In: Baruch A. Brody, ed. Readings in the Philosophy of Science. Englewood Cliffs, N.J.: Prentice-Hall, 1970; pp. 296-306. Reprinted from: Philosophy and Phenomenological Research, I, no. 3 (1951).

_____. The Structure of Science. London: Routledge & Kegan Paul, 1961.

Nowell-Smith, P. H. Ethics. Harmondsworth, Eng.: Penguin Books, 1954.

_____. "Ifs and Cans". In: Bernard Berofsky, ed. Free Will and Determinism. New York: Harper & Row, 1966; pp. 322-339. Reprinted from: Theoria, XXVI, Part 2 (1960), pp. 85-101.

Ofstad, Harold. "Recent Work on the Free-Will Problem". American Philosophical Quarterly, vol. 4, no. 3 (July, 1967), pp. 179-207.

Oppenheim, Paul and Hilary Putnam. "Unity of Science as a Working Hypothesis". In: Herbert Feigl, Michael Scriven, and Grover Maxwell, eds. Minnesota Studies in the Philosophy of Science, vol. II. Minneapolis: University of Minnesota Press, 1958; pp. 3-36.

Pap, Arthur. "Determinism, Freedom, Moral Responsibility, and Causal Talk". In: Sidney Hook, ed. Determinism and Freedom in the Age of Modern Science. New York: New York University Press, 1958; pp. 200-206.

Parsons, Talcott and Edward A. Shils, eds. Toward a General Theory of Action. New York: Harper & Row, 1951.

Pears, D. F. Freedom and the Will. London: Macmillan & Company, Ltd., 1965.

_____. "Two Problems About Reasons for Actions". In: Robert Binkley, Richard Bronaugh, and Ausonio Marras, eds. Agent, Action, and Reason. Toronto: University of Toronto Press, 1971; pp. 128-153. With comments by Irving Thalberg, pp. 154-166.

Peirce, Charles S. "The Doctrine of Necessity Examined". In: Vincent Tomas, ed. Charles S. Peirce: Essays in the Philosophy of Science. Indianapolis: The Bobbs-Merrill Company, 1957. Reprinted from: The Monist, II (1892), pp. 321-337.

Peters, R. S. The Concept of Motivation. London: Routledge & Kegan Paul, 1958.

Polanyi, Michael. Personal Knowledge. New York: Harper & Row, 1958.

Popper, Karl R. "Of Clouds and Clocks". In: Objective Knowledge: An Evolutionary Approach. Oxford: Clarendon Press, 1972; pp. 206-255. Reprinted from Arthur Holly Compton Memorial Lecture, Washington University, 21 April 1965.

_____. "On the Theory of the Objective Mind". In: Objective Knowledge: An Evolutionary Approach. Oxford: Clarendon Press, 1972; pp. 153-190. Reprinted from: Akten des XIV. Internationalen Kongresses für Philosophie, vol. 1, Vienna 1968, pp. 25-53, with some material from Schweizer Monatshefte, 50. Jahr, Heft 3, 1970, pp. 207-215.

Pratt, Carroll C. "Free Will". In: Paul K. Feyerabend and Grover Maxwell, eds. Mind, Matter, and Method. Minneapolis: University of Minnesota Press, 1966.

Quine, Willard Van Orman. "Mr. Strawson on Logical Theory". In: Irving M. Copi and James A. Gould, eds. Contemporary Readings in Logical Theory. New York: Macmillan, 1967; pp. 236-252. Reprinted from: Mind, v. 63 (1953), pp. 433-451.

Radner, Michael. "Popper and Laplace". In: Michael Radner and Stephen Windkur, eds. Minnesota Studies in the Philosophy of Science, vol. IV: Analyses of Theories and Methods of Physics and Psychology. Minneapolis: University of Minnesota Press, 1970; pp. 417-427.

Rankin, K. W. Choice and Chance. Oxford: Basil Blackwell, 1961.

Reid, Thomas. Essays on the Active Powers of the Human Mind. Cambridge: M.I.T. Press, 1969. Reproduced from: The Words of Thomas Reid, vols. III & IV. Charlestown, Mass.: Samuel Etheridge, Jr., 1815.

Rescher, Nicholas. "On the Characterization of Actions". In: Myles Brand, ed. The Nature of Human Action. Glenview, Ill.: Scott, Foresman and Company, 1970; pp. 247-254.

Richman, Robert J. "Responsibility and the Causation of Actions". American Philosophical Quarterly, vol. 6, no. 3 (July 1969), pp. 186-197.

Rollins, C. D. "Are Mental Events Actually Physical?" In: C. F. Presley, ed. The Identity Theory of Mind. St. Lucia, Queensland: University of Queensland Press, 1967; pp. 21-37. Originally presented at First Annual Congress of the Australasian Association of Philosophy, University of Queensland (Brisbane), 1964.

Rosenblueth, Arthuro. Mind and Brain. Cambridge: The M.I.T. Press, 1970.

Rudner, R. Philosophy of Social Science. Englewood Cliffs, N.J.: Prentice-Hall, 1966.

Russell, Bertrand. "On the Notion of Cause". In: Mysticism and Logic. Garden City, N.J.: Doubleday Anchor Books, 1917; pp. 174-201. Reprinted from: Proceedings of the Aristotelian Society, 1912-13.

Russell, Bertrand. "On the Notion of Cause, with Application to the Free-will Problem". In: Herbert Feigl & May Brodbeck, eds. Readings in the Philosophy of Science. New York: Appleton-Century-Crofts, Inc., 1953; pp. 387-407. Reprinted from: Bertrand Russell, Mysticism and Logic. Barnes & Noble, Inc., 1917; pp. 180-205. Also Our Knowledge of the External World. London: W. W. Norton & Co., 1929; pp. 247-256.

Ryle, Gilbert. The Concept of Mind. New York: Barnes & Noble, Inc., 1949.

_____. Dilemmas. Cambridge: The University Press, 1964.

Sartre, Jean-Paul. Being and Nothingness. Hazel E. Barnes, trans. New York: Washington Square Press, 1953.

Scheffler, Israel. The Anatomy of Inquiry. New York: Alfred A. Knopf, 1967.

Schlick, Moritz. "Causality in Everyday Life and in Recent Science". In: Herbert Feigl and Wilfrid Sellars, eds. Readings in Philosophical Analysis. New York: Appleton-Century-Crofts, 1949; pp. 515-533. Reprinted from: University of California Publications in Philosophy, 15 (1932).

_____. Problems of Ethics. David Rynin, trans. New York: Dover Publications, 1939.

Scriven, Michael. "The Key Property of Physical Laws--Inaccuracy". In: Herbert Feigl and Grover Maxwell, eds. Current Issues in the Philosophy of Science. New York: Holt, Rinehart and Winston, 1961; pp. 91-104.

_____. "The Mechanical Concept of Mind". In: Kenneth M. Sayre and Frederick J. Crosson, eds. The Modeling of Mind. New York: Simon and Schuster, 1963; pp. 243-254. Reprinted from: Mind, v. LXII (1953), pp. 230-40.

_____. "A Possible Distinction Between Traditional Scientific Disciplines and the Study of Human Behaviour". In: Herbert Feigl and Michael Scriven, eds. Minnesota Studies in the Philosophy of Science, vol. I. Minneapolis: University of Minnesota Press, 1956; pp. 330-339.

Scriven, Michael. "Truisms as the Grounds for Historical Explanations". In: Patrick Gardiner, ed. Theories of History. Glencoe, Ill.: The Free Press; pp. 443-475.

Searle, John R. "How To Derive 'Ought' from 'Is'". In: W. D. Hudson, ed. The Is-Ought Question. London: Macmillan, 1969; pp. 120-134. Reprinted from: The Philosophical Review, LXXIII (1964).

_____. Speech Acts. Cambridge: University Press, 1969.

Sellars, Wilfrid. "Counterfactuals, Dispositions, and the Causal Modalities". In: Minnesota Studies in the Philosophy of Science, vol. II: Concepts, Theories, and the Mind-Body Problem, ed. by Herbert Feigl, Michael Scriven, and Grover Maxwell. (Minneapolis: University of Minnesota Press, 1958; pp. 225-308).

_____. "Fatalism and Determinism". In: Keith Lehrer, ed. Freedom and Determinism. New York: Random House, 1966; pp. 141-174.

_____. "Notes on Intentionality". In: Wilfrid Sellars, ed. Philosophical Perspectives. Springfield, Ill.: Charles C. Thomas, 1967, pp. 308-320. Reprinted from: Journal of Philosophy, vol. 61 (1964).

_____. "Some Reflections on Language Games". In: Science, Perception and Reality. London: Routledge & Kegan Paul, 1963, pp. 321-358.

_____. "Thought and Action". In: Keith Lehrer, ed. Freedom and Determinism. New York: Random House, 1966; pp. 105-139.

Smart, J. J. C. "Sensations and Brain Processes". In: V. C. Chappell, ed. The Philosophy of Mind. Englewood Cliffs: Prentice-Hall, 1962; pp. 160-172. Reprinted and revised from: The Philosophical Review, LXVIII (1959), pp. 141-156.

Stout, A. K. "Free Will and Responsibility". In: Wilfrid Sellars and John Hospers, eds. Readings in Ethical Theory. New York: Appleton-Century-Crofts, Inc., 1952; pp. 537-548. Reprinted from: Proceedings of the Aristotelian Society, 37 (1936-37).

Stout, G. F. Mind and Matter. Cambridge: University Press, 1931.

Strawson, P. F. Introduction to Logical Theory. London: Methuen, 1952.

_____. "Persons". In: Herbert Feigl, Michael Scriven, and Grover Maxwell, eds. Minnesota Studies in the Philosophy of Science, vol. II. Minneapolis: University of Minnesota Press, 1958; pp. 330-353.

Sumner, L. W. "Value Judgments and Action". Mind, vol. LXXVII, no. 307 (July 1968), pp. 383-399.

Taylor, Charles. "Explaining Action". Inquiry, vol. 13 (1970), pp. 54-89.

_____. The Explanation of Behaviour. London: Routledge & Kegan Paul, 1964.

_____. "The Explanation of Purposive Behaviour". In: Robert Borger and Frank Cioffi, eds. Explanation in the Behavioural Sciences. Cambridge: University Press, 1970; pp. 49-79. With comments by Robert Borger, pp. 80-88, and reply by Charles Taylor, pp. 89-95.

_____. "How is Mechanism Conceivable?" In: Marjorie Grene, ed. Interpretations of Life and Mind. London: Routledge & Kegan Paul, pp. 38-64. Reprinted with revisions from: A. Karazmer and J. C. Eccles, eds. Brain and Human-Behaviour.

_____. "Mind-Body Identity, A Side Issue?" The Philosophical Review, vol. LXXVI, no. 2 (April, 1967), pp. 201-213.

Taylor, Richard. Action and Purpose. Prentice-Hall, Inc.: Englewood Cliffs, N.J.; 1966.

_____. "Determinism and the Theory of Agency". In: Sidney Hook, ed. Determinism and Freedom in the Age of Modern Science. New York: New York University Press, 1958; pp. 211-218.

_____. "Prevention, Postvention, Will". In: Keith Lehrer, ed. Freedom and Determinism. New York: Random House, 1966; pp. 65-85.

_____. "Thought and Purpose". In: Myles Brand, ed. The Nature of Human Action. Glenview, Ill.: Scott, Foresman and Company, 1970; pp. 267-282. Reprinted from: Inquiry, II, no. 2 (1969).

Toulmin, Stephen. The Philosophy of Science. London: Hutchinson University Library, 1953.

_____. "Reasons and Causes". In: Robert Borger and Frank Cioffi, eds. Explanation in the Behavioural Sciences. Cambridge: University Press, 1970; pp. 1-26. With Comment by R. S. Peters, pp. 27-41, and Reply by Stephen Toulmin, pp. 42-48.

University of California Associates. "The Freedom of the Will". In: Herbert Feigl and Wilfrid Sellars, eds. Readings in Philosophical Analysis. New York: Appleton-Century-Crofts, Inc., 1949; pp. 594-615. Reprinted from: University of California Associates. Knowledge and Society. New York: D. Appleton-Century Company, 1938.

Warnock, G. J. "'Every Event Has a Cause'". In: Antony Flew, ed. Logic and Language (second series). Oxford: Basil Blackwell, 1961; pp. 95-111.

Weiss, Paul. "Common Sense and Beyond". In: Sidney Hook, ed. Determinism and Freedom in the Age of Modern Science. New York: New York University Press, 1958; pp. 218-224.

White, Morton. Foundations of Historical Knowledge. New York: Harper & Row, 1965.

Winch, Peter. The Idea of a Social Science. London: Routledge & Kegan Paul, 1958.

Yolton, John W. "Agent Causality". American Philosophical Quarterly, vol. 3, no. 1 (January 1966), pp. 14-26.